Function of miRNAs in murine liver carcinogenesis

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von

M.Sc. Ivana Winkler

aus Zagreb, Kroatien

Tübingen
2019

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

Tag der mündlichen Qualifikation:	17.10.2019
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Alfred Nordheim
2. Berichterstatter:	Prof. Dr. Boris Macek
3. Berichterstatter:	Prof. Dr. Peter Angel

Every passing hour brings the Solar System forty-three thousand miles closer to Globular Cluster M13 in Hercules - and still there are some misfits who insist that there is no such thing as progress.

Kurt Vonnegut Jr., The Sirens of Titan

Declaration

All work presented here was performed by Ivana Winkler, unless stated otherwise. Colleagues contribution is listed in details in **Appendix C: Contributions**. Thesis itself, was written by Ivana Winkler. The work was performed under supervision of Prof. Alfred Nordheim during the period from September 2014-December 2018 at the Interfaculty Institute of Cell Biology, University of Tübingen.

Abstract

Hepatocellular carcinoma (HCC) is the most common primary liver malignancy. Its progression is a multistage process that typically arises in the context of liver fibrosis. Liver fibrosis is a consequence of an exaggerated wound healing response to reoccurring or chronic liver injury, characterized by excessive accumulation of ECM. This process ultimately results in scarring and thickening of affected tissue, which interferes with normal liver function and facilitates HCC tumorigenesis. Although, liver fibrosis and HCC pose a major threat to human health, regulatory networks governing the events leading to their development are insufficiently understood. Accumulating data support a regulatory role of microRNAs (miRNAs) in control of gene expression programs that underline different normal and pathologic processes, including cancer. In this study, the *SRF-VP16*^{iHep}HCC model was used as starting point to investigate the role of miRNAs in regulation of HCC and associated microenvironment development. The resulting analysis identified a miRNA network composed of 8 miRNA hubs and 54 target genes. Collectively, let-7 and miR-30 families, miR-29c, miR-335 and miR-338 (termed anti-fibrotic microRNAs, AF-miRNAs) downregulate key structural, signaling and remodeling components of the extracellular matrix. During fibrogenic transition, these miRNAs are transcriptionally regulated by the transcription factors Pparg and Egr1. Thus, this study identified a new role of Pparg and Egr1 as regulator of a functionally related class of AF-miRNAs. The miRNA network is active in human HCC, breast and lung carcinomas, as well as in two independent mouse fibrosis models. Therefore, identified miRNA:mRNA network regulates fibrosis of tumorous and non-tumorous organs, the liver in particular, in mice and humans.

Zusammenfassung

Das hepatozelluläre Karzinom (HCC; aus dem Englischen: Hepatocellular carcinoma) ist der häufigste primäre Lebertumor. Die Entwicklung des Tumors ist ein mehrstufiger Prozess im Kontext der Leberfibrose. Letztere, charakterisiert durch ekzessive Anhäufung von Komponenten der extrazellulären Matrix (ECM; aus dem Englischen: Extracellular matrix), ist Folge einer verstärkten Wundheilungsreaktion bei häufigen oder chronischen Leberschädigungen. Das Resultat ist eine Vernarbung der Leber und eine Verdickung des betroffenen Gewebes, was wiederum zu einer Abweichung von der normalen Leberfunktion führt und die Entstehung von HCC begünstigt. HCC und Leberfibrose sind enorme Bedrohungen für die menschliche Gesundheit. Allerdings sind die regulatorischen Netzwerke, die ihre Entstehung und Entwicklung beeinflussen und bestimmen nur unzureichend erforscht und verstanden. Forschungsdaten legen eine wichtige Rolle für microRNAs (miRNAs) bei der Kontrolle der Genexpression im Allgemeinen und der Kontrolle der meisten pathologischen Prozesse, dem Krebsleiden im Speziellen, nahe. In der vorliegenden Arbeit wird das *SRF-VP16*^{iHep}HCC Model (M. musculus) als Ausgangspunkt für eine Untersuchung der miRNA-abhängigen Kontrolle der extrazellulären Matrix und seiner Rolle im HCC Kontext benutzt. Das Resultat der Analyse ist die Identifikation eines miRNA-Netzwerkes welches aus 8 sogenannten Hub-Knoten und 54 Zielgenen besteht. Die let-7 und miR-30 miRNA-Familien, zusammen mit den miRNAs miR-29c, miR-335 und miR-338 (im nachfolgenden als anti-fibrotische miRNAs, AF-miRNAs, bezeichnet) unterdrücken die Expression von entscheidenden Struktur-, Signal- und Remodellierungskomponenten der extrazellulären Matrix. Bei Entstehung der Fibrose werden die AF-miRNAs durch die Transkriptionsfaktoren Pparg und Egr1 transkriptionell reguliert. Die Arbeit zeigt demnach die Rolle dieser beiden Transkriptionsfaktoren als Regulatoren von miRNAs mit überlappender und konsistenter Funktionalität. Darüber hinaus ist das miRNA Netzwerk auch im menschlichen HCC, Brust- und Lungenkrebs und zwei weiteren Mausmodellen für Leberfibrose aktiv. Zusammenfassend wird gezeigt, dass das identifizierte miRNA:mRNA-Netzwerk

zur Regulierung der Fibrose in Maus und Mensch in tumorösen und nicht-tumorösen Organen, der Leber im Speziellen, beiträgt.

Acknowledgments

I would like to thank Alfred for his supervision and guidance during my PhD process. I would like to thank as well my TAC committee members, Boris and Elisa for their support and valuable advice.

Regrettably, on 30th of April 2018, Eliza Izaurralde passed away. I would like to use this opportunity to once more express all my gratitude to her.

My further thanks goes to all members of Nordheim lab especially:

- Heidemie and Catrin for their big practical contribution to the experimental work and very essential small discussion over coffee breaks in the office;
- Michael for all discussions and his support after all failed experiments,
- Elena and Anke for all technical support,
- Heide for all the help with elaborate paperwork;
- Siggi for his advices and help with organisation of my experiments (i.e. introducing me with half of the institute in search of electroporator device);
- Abhi for his help with mouse work and initial orientation when I started my PhD work;
- Catrin, Michael and Mohan for all the nice lunches we had together
- all my students: Nina, Mario, Katrin, Sandra, Catrin, Franziska

I would like to thank, as well, our collaborators: Dr. Robert Geffers, Dr. Sabin Bhuju, Dr. Katarina Matic, Prof. Dr. Boris Macek, Prof. Dr. Ralf Weiskirchen, Prof. Dr. Jan Hengstler, Brigitte Begher-Tibbe, Dieter Weichenhan, Marion Bähr, Dr. Erawan Borkham-Kamphors and Sebastian Winkler.

My special thanks goes to my family: my parents and siblings for all of their support. My biggest thanks goes to my husband Sebastian, without whom this all would be pointless.

Contents

Chapter 1

Introduction

1.1 Hepatocelular carcinoma

1.1.1 Incidence and risk factors

Hepatocellular carcinoma (HCC) is the most common primary liver malignancy, accounting for 90% of cases[1]. With more than 850,000 new cases annually worldwide HCC represents fifth most common cancer in men and the seventh in women[2]. This neoplasm is currently the third leading cause of cancer-related death[3,4].

The development of HCC is closely related to chronic formation and progression of fibrosis and cirrhosis. Fibrosis is defined as an excessive accumulation of connective and scarring tissue in an organ during regenerative processes while cirrhosis is considered as a late stage of fibrosis which involves significant liver damage caused by inflammation and scarring[2]. Most of the risk factors for HCC development, therefore, lead to chronic liver disease caused by re-occurring liver damage and regeneration cycles. Major risk factors include: hepatitis B virus (HBV) or hepatitis C virus (HCV) infection, alcoholic liver disease and metabolic syndrome[5]. Other co-factors such as tobacco inhalation and exposure to aflatoxin B1 are well-characterised contributors to HCC[1].

Distribution of these risk factors among HCC patients is variable and primarily depends on geographical region[2] (Figure 1.1). Most of the incidences of HCC (85%) occur in developing countries of Southeast Asia and sub-Saharan Africa, in the regions with low-income and sparse resources where HBV is endemic. As HBV infection typically is contracted at an early age, affected individuals develop HCC in their mid-adulthood,

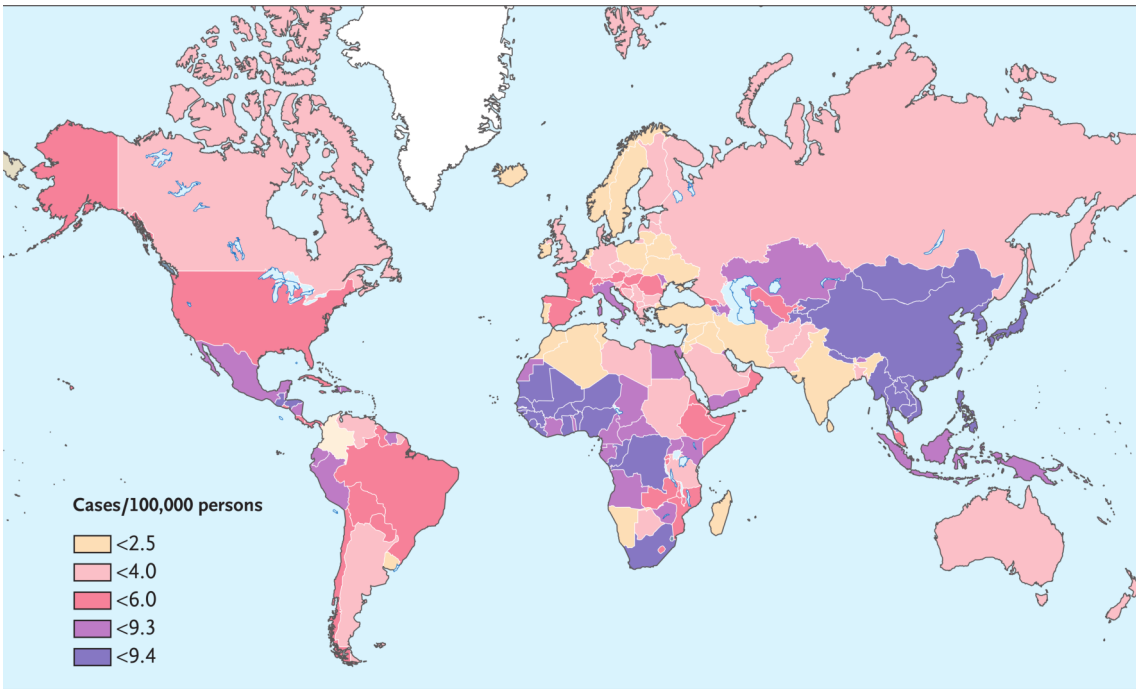

Figure 1.1: The incidence rates of HCC shown as number of cases per 100,000 persons irrespective of age and sex. Reprinted from El-Serag, 2011[2].

during the most productive years of their lives[4]. In contrast, in developed countries (North America, Europe and Japan) HCV is the leading cause of HCC. Additional risks in these countries include excessive alcohol use and non-alcoholic staetohepatatis (NASH) linked to metabolic syndrome. While the contribution of HCV infections to HCC occurrence is expected to decline in the near future, the increasing prevalence of the metabolic syndrome, diabetes mellitus and NASH are all expected to contribute to increased rates of HCC in developed countries[4].

1.1.2 Diagnosis and staging

At the stage when HCC becomes symptomatic, disease has already progressed to advanced stage and it is not longer considered curable. However, HCC has a prolong course which allows for early detection. If diagnosed at an very early stage, HCC lesions are small and potentially curative[3]. Therefore, it is recommended that patients with heightened risk levels of HCC development (e.g. patients infected with HBV or HCV and patients manifesting alcoholic and non-alcoholic fatty liver disease) are enrolled in surveillance programs[3]. Patients who are part of the surveillance protocols

usually are diagnosed by identification of new dysplastic liver nodules on abdominal ultrasound in the asymptomatic stage of HCC and their diagnosis is confirmed by biopsy or other non-invasive criteria. Predictably, patients outside the surveillance programs usually present advance stage of HCC with large symptomatic tumors. Symptoms include weight loss, anorexia, abdominal pain and symptoms related to severe liver dysfunction[1].

Multiple staging systems for HCC classification have been described, including: Barcelona Clinic Liver Cancer (BCLC), Cancer of Liver Italian Program, TNM (Tumor, node and metastasis), Okuda and Japanese Integrated Staging Score systems. BCLC system is currently the only system endorsed by the American Association for the Study of Liver Diseases and the European Association for the Study of the Liver[4]. Additionally, BCLC staging has been proposed as the standard for use in clinical practice as it has an advantage that tumor staging is linked with treatment recommendations[4].

The BLCL system incorporates following set of parameters in order to assess tumor stage: patient's performance status, number and size of nodules, tumor related symptoms and liver function assessment as determined by Child-Pugh classification system[2]. The Child-Pugh scoring system is defined by five clinical measurements of liver diseases. Each measurement is assigned a score on a scale from 1 to 3 and score are summed up over five measurements. A final score of 5 or 6 points indicates class A disease (well-preserved liver function), 7 to 9 class B (moderate liver dysfunction with life expectancy of 3 y) and 10 to 15 points class C (most severe disease with life expectancy of 1 y)[2].

The BLCL system distinguishes five different stages of HCC: very-early-stage, early-stage, intermediate-stage, advance-stage and terminal-stage[3].

Very-early-stage is currently difficult to diagnose as it manifests as single, asymptomatic lesion less than 2 cm in diameter of size, with no vascular or distant metastasis[2].

Early-stage HCC underlines some preserved liver function (Child-Pugh class A or B) with a single nodule measuring less than 5 cm or no more than 3 nodules each with less than 3 cm in diameter[2].

Patients with manageable cirrhosis, no symptoms and no vascular lesions, but with large or multi-focal lesions are considered to have intermediate-stage HCC[2].

Patients with mild HCC-related symptoms, vascular invasion or extra-hepatic metastasis are considered to be in advanced-stage of HCC[2].

1.1.3 Treatment

If HCC is diagnosed at very-early-stage or early-stage patients are eligible for curative therapy: resection, liver transplantation or percutaneous local ablative therapy. Treatment in these cases is associated with overall survival rate of 90%[4].

Surgical resection is the method of choice for patients without cirrhosis, with a single resectable HCC and with preserved liver function. However, 5-year risk of recurrence exceeds 70% as patient's remaining liver continues to be at risk of spread of interliver metastasis and *de novo* tumor development caused by the primary source of the liver injury[3].

Ablation promotes tumor necrosis by thermal injury or injection of chemical agents. It has close to 100% efficiency in removing HCC nodules that are less than 2 cm in diameter. Survival rates after ablation therapy are almost identical to survival rates following resection[3].

Orthopic liver transplantation is the only definite treatment of the HCC since it removes malignancy together with the diseased liver and thus eliminates the high risk of HCC re-occurrence[4]. However, as this procedure is very resource intensive, strict criteria are applied to select appropriate candidates. Candidates are eligible for a liver transplant if they fulfill Milan criteria. Milan criteria dictate that a suitable candidate should have tumor lesion less than 5 cm in diameter or up to three lesions with the largest no more than 3 cm in diameter. Patients which undergo liver transplantation have an expected 4-year survival rate of 85% and a recurence-free survival rate of 92%[2].

Patients diagnosed with intermediate-stage typically undergo transarterial chemoembolisation (TACE) therapy. TACE is a minimally invasive procedure performed to restrict tumor blood supply. It combines local delivery of beads coated with a chemotherapeutic drug injected through the catheter to an artery with local administration of chemotherapy. This procedure improves the 2-year survival by 20-25% as compared with conservative therapy[3].

The primary treatment option for patients with an advanced-stage of HCC is the oral agent sorafenib. This drug blocks the activity of Raf ser/thr kinase, together with Vascular Endothelial Growth Factor Receptors (VEGFR) and Platelet-derived growth factor receptor (PDGFR) signaling pathways, thus targeting tumor angiogenesis and tumor cell proliferation[5]. The 1-year survival rate for patients with advanced-stage is less than 10%[2]. Systemic chemotherapy with doxorubicin, PIAF (platinum, interferon,

doxorubicin, 5-FU), or FOLFOX (folinic acid, 5-FU and oxaliplatin) does not increase survival rates and was accompanied with significant toxicity in patients with significant liver damage[1].

In summary, to obtain the best treatment option for HCC, early diagnosis and early treatment start are crucial. Therefore, potent liver-directed anti-fibrotic therapies focused on targeting of pre-cancerogenous stage might reduce the risk of HCC development[1].

1.1.4 Molecular pathogenesis and pathophysiology

HCC development is a multi-step process that typically occurs in the setting of liver cirrhosis. HCC development starts with the appearance of pre-cancerous cirrhotic nodules with low-grade dysplasia. Low-grade nodules gradually develop into high-grade dysplastic nodules, some of which can then transform into early-stage HCC[1].

Similar to most solid tumors, HCC is the result of the accumulation of different mutations in passenger and driver genes accompanied with epigenetic mutations.

Most frequent somatic mutation in HCC is Telomerase reverse-transcriptase (*TERT*) promoter mutation. Additionally, this mutation has been identified as the earliest genomic event in the multistep process of HCC development[6]. *TERT* promoter mutation correlates with increased *TERT* expression and thus influences telomere maintance[7] (Figure 1.2).

β-catenin (*CTNNB1*) is another frequently mutated gene in HCC. Functional consequence of *CTNNB1* mutation is a higher activity of the WNT pathway[8]. In addition to *CTNNB1*, other mutations, which lead to higher activity of WNT pathway, were observed in WNT regulators, axin 1 (*AXIN1*) and adenomatous polyposis coli (*APC*)[9].

A large number of genes encoding chromatin remodelers or transcriptional regulators are repeatedly modulated in HCC. This includes BRG1- or HRBM-associated factors (BAFs) and polybromo-associated BAF chromatin complexes. Additionally, mutations often occur in histone methlyation writer family (KMT2) leading to alterations in chromatin state and altered gene expression[1].

Lastly, RAS-RAF-MAPK and the phosphoinositide 3-kinase-AKT-mTOR pathways are frequently activated in HCC[9].

Genomics analysis conducted over the past decades led to better understanding of molecular pathogenesis of HCC. However, despite the new insights into the molecular

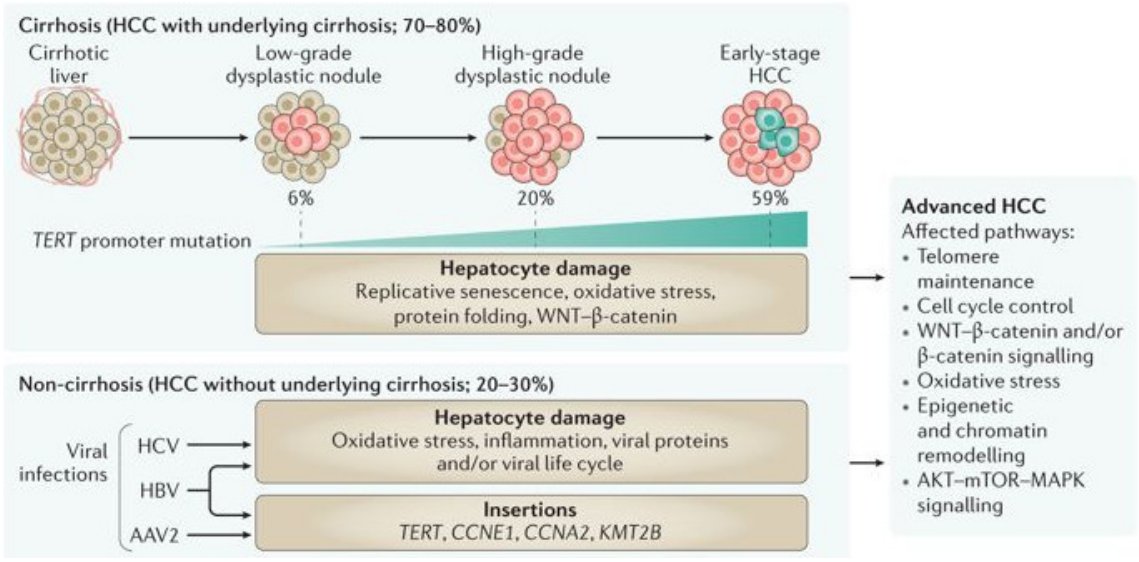

Figure 1.2: Most frequent molecular events observed in liver carcinogenesis. Most frequent and earliest genomic mutation in HCC development is *TERT* promoter mutation. Another mechanism that promotes HCC formation is the activation of an oncogene as a consequence of adeno-associated virus 2 or hepatitis C and B virus infection. Reprinted and modified from Llovet et al., 2016[1].

pathogenesis of the HCC, none of the molecular classifications that have been proposed predicted accurately the disease outcome or recurrence[3].

1.2 Tumor microenvironment

The tumor microenvironment is a complex composite of tumor cells embedded within the extracellular matrix (ECM) and combined with a mix of different stromal cells and their secreted proteins. An altered tumor microenvironment is a key component of carcinogenesis and it is considered to be involved in all stages of malignant progression, from transformation to invasion and metastasis[10]. Research conducted so far, clearly indicates a contribution of stromal cells to all hallmarks of cancer: sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, activating invasion and metastasis, reprogramming energy metabolism and evading immune destruction[11]. Therefore, targeting components of the tumor microenvironment to nullify their tumor-supportive role represent a tempting therapeutic strategy.

The development of HCC is a multi-stage process usually initiated by a chronic insult (e.g. HCV, HBV infection, alcohol) to the liver. The chronic insult causes liver injury through reactive oxygen species production, cellular DNA damage, endoplasmatic reticulum stress and necrosis of damaged hepatocytes. The following injury response of the liver involves the activation of hepatic stellate cells (HSC) and macrophages, which produce components of the ECM and growth factors essential for migration of endothelial cells, neoangiogenesis, fibrosis and inflammation. The underlying process in the context of inflammation and oxidative DNA damage promotes development of dysplastic nodules and their malignant transformation to HCC[10](Figure 1.3).

1.2.1 Angiogenesis

HCC is a prototypical highly vascularized tumor and angiogenesis plays a major role in the carcinogenesis from its early stages[10].

The hepatic wound-healing response triggered by liver injury leads to fibrogenesis. In turn, the prolonged fibrogenic process leads to an abnormal angioarchitecture. Formation of fibrotic septa provokes resistance to blood flow resulting in reduced metabolic exchange of oxygen, causing hypoxia. Hypoxia induces transcription of hypoxia-sensitive genes which in turn stimulate the formation of new vessels[12].

Once a tumor emerges and grows, the survival of neoplastic cells requires the formation of a new vasculature in order to provide nutrients and oxygen. Growth of the tumor creates a nutrient and oxygen deprived environment which leads to activation of endothelial cell proliferation. The neoplastic cells and other components of the tumor microenvironment, primarily HSC and macrophages, secrete a number of angiogenic growth factors. The most critical pro-angiogenic factor is vascular endothelial growth factor (VEGF). VEGF binds to its receptors VEGFR1 and VEGFR2 and activates signaling pathways involved in proliferation, migration and invasion of endothelial cells. The expression of VEGF correlates with poor prognosis in HCC[10].

1.2.2 Chronic inflammation

HCC represents an inflammation-associated tumor, densely infiltrated by the innate and adaptive immune cells[1].

Several inflammatory mediators have been determined to participate in sustained inflammation and immuno-suppression associated with development of HCC.

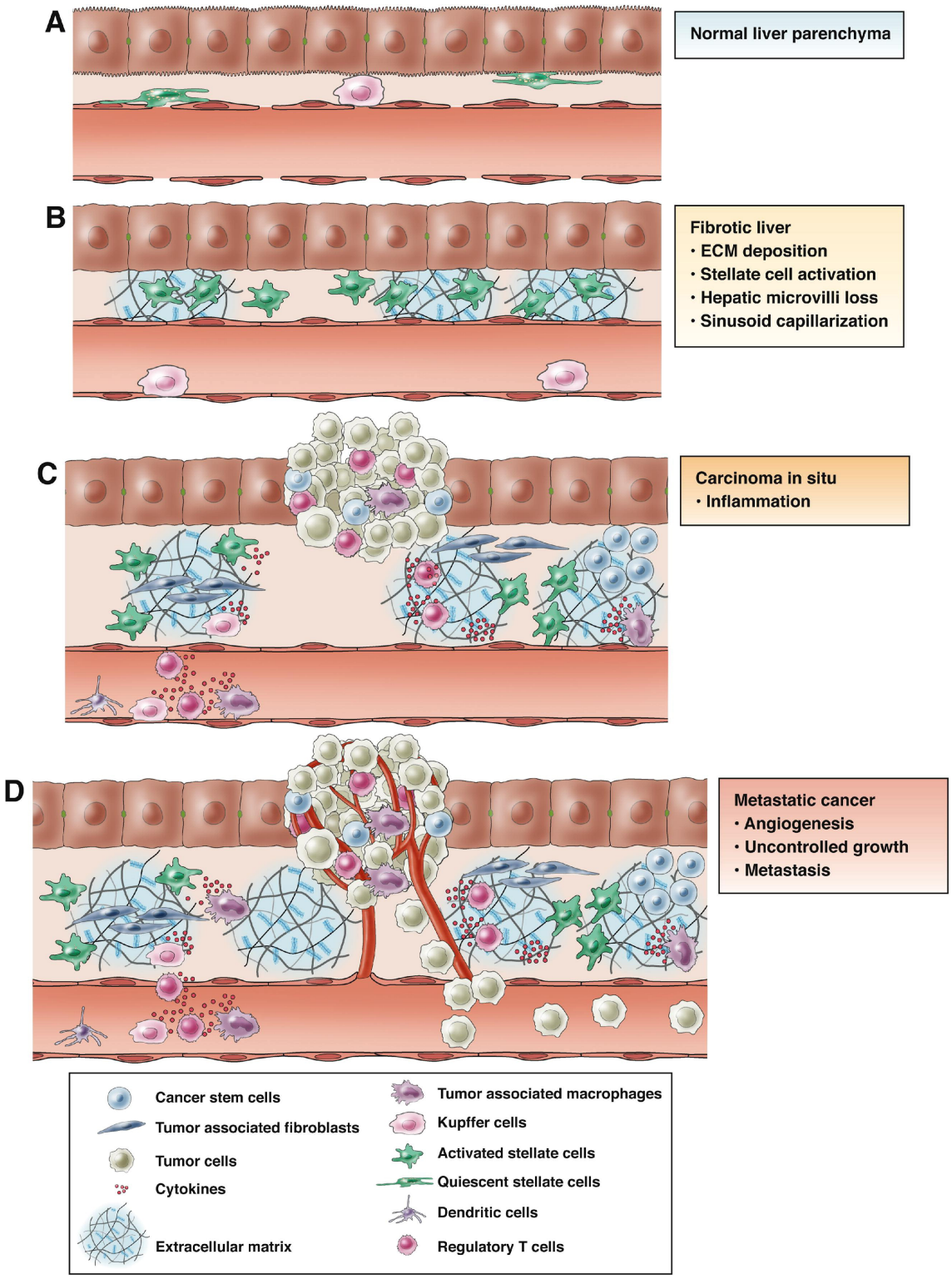

Figure 1.3: Cellular alterations leading to HCC development. A) Normal liver parenhyma. B) Liver injury leads to activation of hepatic stellate cells and macrophages, which produce components of the ECM and growth factors essential for migration of endothelial cells, neoangiogenesis, fibrosis and inflammation which stimulate development of HCC (C). D) Development of new blood vessels and distant metastasis. Reprinted from Hernandez-Gea et al, 2013[10].

Continuous cytokine production can stimulate many liver cell types that are associated with carcinogenesis. High levels of interleukin (IL)-6, IL-10 and IL-22 are associated with high risk of HCC development and poor prognosis[10].

Nuclear factor κB (NF-κB) and STAT3 are signaling pathways involved in the hepatic inflammatory response to injury that is critical for liver regeneration and promotion of HCC in the inflamed liver[10].

Several growth factors, primarily transforming growth factor (TGF)-β, hepatocyte growth factor (HGF) and epidermal growth factor (EGF), regulate immune response in HCC microenvironment[10].

1.2.3 Fibrosis

Hepatic fibrosis is an essential part of the wound-healing response characterized by the accumulation of ECM. If the insult to the liver is acute, ECM changes are transient and reversible. However, if the injury is chronic, accumulation of ECM and accompanying inflammation persist, leading to substitution of parenchyma with scar tissue. This results in cirrhosis, the end consequence of the fibrosis[13].

The hepatic parenchyma is composed of epithelial cells (hepatocytes), endothelial cells, and resident non-parenchyma cells, mainly hepatic stellate cells (HSCs) and Kuppfer cells. HSCs reside in the sub-endothelial space of Disse, located between the hepatocyte and sinusoidal endothelial lining[14].

Following liver injury, HSCs become activated which leads to dramatic change of their phenotype. Most distinctive feature of quiescent HSCs is the storage of vitamin A (retinol) in the form of retinyl esters in cytoplasmic lipid droplets[15]. One of the defining characteristic of HSC activation is loss of retinol droplets followed by enhanced proliferation, contraction and release of pro-inflammatory, pro-fibrotic and pro-mitogenic cytokines[16].

HSC activation is divided in two phases: initiation and perpetuation. Initiation involves early changes in gene expression and phenotype and it is a result of paracrine stimulation from damaged hepatocytes. Rapid induction of growth factor receptor expression, development of contractile and fibrogenic phenotype are principal features of initiation. Perpetuation induced by paracrine and autocrine signaling is characterized by amplification of the activated phenotype[15].

Activated HSCs migrate and accumulate at the site of tissue repair where they secrete large amounts of ECM components. Quantitative and qualitative changes in

ECM deposition result in fibrosis. In advanced stages, the fibrotic liver contains six times more ECM than normal liver. Fibrotic ECM is mostly composed of collagen I and III, fibronectin, elastin, laminin and proteogycans[17].

Reduction of ECM degradation is primarily a result of overexpression of Tissue inhibitor of metalloproteinases (TIMPs). Their overexpression results in decreased activity of ECM-removing metalloproteinases (MMP) and thus, leads to decreased ECM removal[18].

Accumulation of ECM is a key feature of the tumor microenvironment. ECM can modulate the activation and proliferation of HSCs, angiogenesis, and the availability of growth factors and MMPs. Additionally, the ECM provides cells with signals for polarization, adhesion, migration, survival and differentiation[13].

A fibrogenic microenvironment, characterized by increased stiffness and plethora of growth factors, contributes directly to tumorigenesis. Studies involving ultrasound elastography demonstrated that liver stiffness measurement is a strong predictor of HCC development[19]. Matrix stiffness promotes tumorigenesis by increasing integrin signaling, which in turn leads to enhanced growth, survival and proliferation. Integrins are heterodimeric transmembrane receptors with globular head domain which can bind components of the EMC and cell adhesion molecules. Specifically, integrins $\alpha 1\beta 1$ and $\alpha 2\beta 1$ have been implicated in progression and cancer cell invasion[10]. Furthermore, ECM deposition enhances HCC chemotherapy resistance and offers protection against cells of the immune system[12,20].

Inflammatory cytokines play a key role in fibrosis. TGF-β is considered to be the most potent fibrogenic cytokine. TGF-β binding and phosphorylation of type I receptor induces phosphorylation of downstream SMADs, primarily SMAD3[21]. Activation of SMAD3 leads to increased transcription of collagen I and III. Additionally, TGF-β activates mitogen-activated protein kinase (MAPK) signaling to promote further HSC activation[22].

Platelet derived growth factor (PDGF) is a critical mitogen in liver which induces proliferation of HSCs through extracellular signal-regulated kinase (ERK)-dependent and ERK-independent mechanisms[17]. PDGF is a dimeric protein composed of combinations of four polypeptide chains (A, B, C and D). All PDGF isoforms are upregulated during HSC activation and correlate with the degree of fibrosis[13].

Epigenetic regulation plays an essential role in HSCs activation. Epigenetic modifications consist of three main modes of regulation: histone modifications, DNA methylation and silencing by non-coding RNAs.

Trichostatin A, a histone deacetylase inhibitor, blocks the phenotypical features of HSCs and reduces both HSC proliferation and transcription of *collagen I*[13]. Myocardin-related transcription factor A (MRTF-A) stimulates pro-fibrotic transcription by recruiting a histone metlytransferase to the promoters of fibrogenic genes[15].

Methylation of cytosines at cytosine-phospho guanine (CpG) dinucleotide positions in promoter regions represses gene transcription. Methyl-CpG binding protein 2 (MECP2), which specifically binds methylated DNA, represses anti-fibrotic nuclear receptor peroxisome proliferator-activated receptor γ (PPARγ) to promote HSC activation[15].

Among epigenetic signals, microRNAs (miRNAs) play a significant role in regulation of HSC activation and fibrosis development. Implicated miRNAs include miRNA-21, miRNA-133a, miRNA-122, miRNA-214, miRNA-221/222, miRNA-29b and others[23].

HSCs express diverse nuclear transcription factor receptors. Specific ligand binding in the cytoplasm activates nuclear receptors and causes their translocation to the nucleus where they can regulate the expression of the target genes.

Although, retinoid storage is one of the more distinguishing features of HSCs, the relative importance of retinoic acid receptors (RARs) and retinoid x receptors (RXRs) is not very well understood. However, it has been shown that expression of RAR and RXR decreases with HSC activation[13].

Vitamin D receptor (VDR) ligands inhibit HSCs activation mediated by TGF-β. Additionally, *Vdr*-knockout mice develop spontaneously hepatic fibrosis[15].

The Peroxisome proliferator-activated receptors (PPARs) regulate lipid and glucose metabolism. The most prominent isoform PPARγ is considered to an play important role in the maintenance of HSCs quiescence.

While, in the past, fibrosis was thought to be irreversible, a growing body of evidence shows that under some circumstances fibrosis, even cirrhosis, is reversible. Animal models have demonstrated that removing the source of liver injury results in fibrosis resolution. However, human fibrotic diseases are often multi-factorial and eradication of the injurious stimuli may not always be possible[24].

Effective resolution of fibrosis requires three critical events: elimination of the cause of injury, removal of the fibrotic ECM and elimination of activated HSCs. Intact fibrillar collagen, which is the main component of ECM, can be degraded by collagenases. MMP1, MMP13 and MMP14 are the crucial collagenases necessary for matrix degradation. Alternatively, macrophages and myofibroblast can internalize collagen for degradation[17].

Elimination of activated stellate cells is possible through three alternative cell fates: apoptosis, senescence and de-differentiation. Apoptosis of HSCs is coordinated by extrinsic pathways triggered by death receptors and intristics pathway mediated through mitochondria. However, activated HSCs are often resistant to apoptosis due to upregulation of anti-apoptotic proteins B-cell lymphoma 2 (BCL-2) and Heat Shock Protein Family A (Hsp70) Member 1A/B (HSP1A/B). Additionally, ECM stiffness increases HSC resistance to apoptosis[13].

When undergoing senescence, activated HSCs switch from proliferative, ECM producing cells to growth-arrested, ECM-degrading anti-fibrotic cells which are then removed by natural killer cells or macrophages[24].

Finally, activated HSCs can revert to an inactive quiescent-like phenotype. While de-differentiated HSCs downregulate most of the fibrotic genes (e.g. *Acta2, Col1a1, Col1a2, Timp1* and *Tgfbr1*) some quiescence-associated genes remain inactive (e.g. *Adfp, Adipor1* or *GFAP*. Additionally, de-differentiated HSCs are more responsive to further fibrogenic stimuli than naive HSCs[25].

Considering the clinical relevance of fibrosis and findings that it is reversible, several therapeutic approaches are being developed to combat this serious health issue. Most therapies are focused on: elimination of primary disease, downregulation of early HSC activation, inhibition of HSC proliferation and fibrogenesis, promotion of HSC apoptosis or quiescence and stimulation of ECM degradation[17].

1.3 MicroRNAs

MiRNAs are 21-23 nucleotide long RNAs that act as important regulators of gene expression[26]. Mature miRNAs associate with Argonaute (Ago) protein and together form the RNA-induced silencing complex (RISC), a ribonucleoprotein complex effecting posttranscriptional gene silencing.

Complementary base pairing of the miRNA with target messenger RNA (mRNA) serves as a guide for Ago protein and directs degradation, destabilization or translational repression[27].

In plants, perfect pairing between an miRNA and its target leads to endonucleolytic cleavage by Ago resulting in rapid mRNA degradation[28].

In contrast, metazoan miRNAs typically require pairing of only 6-8 nucleotides at the 5' end (nucleotides 2-8) of the miRNA to successfully recognize and repress

their targets. This region which nucleates miRNA:target mRNA pairing is called the miRNA "seed" region[27]. Pairing of miRNA nucleotides 13-16 to the untranslated region (UTR) region can augment seed paring. Such "3-supplementary sites" are atypical and only slightly more effective compared to canonical "only-seed" sites[29]. Pairing to the 3' portion of the miRNA can, as well, compensate for a single-nucleotide bulge or mismatch in the seed region. These sites are called "3'-compensatory sites". However, compensatory sites are rarely used for biological targeting, comprising only 1 % of the conserved sites in mammals[27].

Partial pairing of the miRNA to 3'-UTR of the target can result in deadenylation induced by the CCR4-NOT complex[30]. Alternatively, partially paired miRNA can induce translational repression by blocking initiation via CCR4-NOT complex recruitment or by inhibiting translational elongation through promotion of ribosome drop-off[28]. Experimental studies, conducted so far, indicate that miRNA-depended changes are enacted largely through mRNA destabilization[31].

miRNA influence on target downregulation can range from a fine-tuning effect, to moderate and substantial repression of protein output. The effect of repression is dependent on different expression levels and profiles of miRNAs in different cell types and expression levels of their target in the same cells[31].

Matching of the miRNA seed region to the 3'-UTR of mRNA can explain much the of targeting specificity, however, *in vivo* target selection is depended on additional parameters. A feature that correlates with repression efficiency is the location of the target site in UTR region. MiRNA binding site are frequently depleted within the 3'-UTR at least 15 nucleotides from stop codon due to translating ribosomes which tend to hinder stable RISC-transcript association and lead to inefficient target repression[31]. Additionally, functional targeting sites are usually positioned away from the center of long UTRs. Presumably, positioning silencing complex in the middle of long UTR could lead to obstructive interactions from complexes bound at either ends of translating mRNA[27]. Another important determinant of miRNA regulation is nucleotide composition near the targeting site. AU-richness and areas devoid of secondary structures enhance accessibility to the miRNA complex[29]. Proximity of sites for co-expressed miRNAs can further promote miRNA silencing. Two sites that are close together (within 40 nt, but further apart then 8 nt) can co-cooperatively repress their target, leading to significantly enhanced target repression[27].

In mammals, more than 60 % of protein-coding genes are believed to be under the control of miRNAs[32]. Functional studies place almost every cellular process investi-

gated to date under miRNA influence and aberrant miRNA expression contributes to a range of human pathologies, including cancer[33]. However, in the majority of cases which characterize the regulatory role of miRNAs, the magnitude of the described miRNA:mRNA regulation is mild. This seeming discrepancy of the extensive role that miRNAs have in different biological processes and the mild extend of their influence is explained by the capacity of individual miRNAs to target hundreds of different mRNAs simultaneously. If the miRNA targets are enriched in common pathways, then the sum of modest effects of individual miRNA:mRNA interactions can produce a stronger response than the direct interactions in isolation[34]. With the potential to target hundreds of mRNAs simultaneously, miRNA themselves have to be tightly and dynamically regulated. miRNA stability and efficiency is controlled by regulatory pathways that affect miRNA biogenesis, processing and alternate mature miRNA target specificity[35,36].

1.3.1 MicroRNA biogenesis

miRNAs are transcribed by RNA polymerase II or RNA polymerase III into primary miRNA (pri-miRNA) (Figure 1.4). Several miRNA-encoding genes are present as clusters in the genome and probably derive from common transcripts[35]. Alternatively, they may be transcribed from individual loci or may originate from exons or introns of other protein-coding genes. miRNAs that reside in exons or introns of protein coding genes share the promoter with the host gene[37].

The pri-miRNAs fold into hairpins, which consist of a stem of 33-35 bp, terminal loop and single-stranded segments at both ends. Such structure is recognized by Drosha, an RNase III family enzyme. Drosha forms together with its essential cofactor DGCR8 a complex called Microprocessor. Microprocessor cleaves pri-miRNA, producing a 70-nucleotide long precursor miRNA (pre-miRNA)[32].

The product of Drosha cleavage is exported to the cytoplasm by exportin 5, where it is processed by Dicer, another RNase III family enzyme[36]. Dicer cleaves pre-miRNa at a fixed distance from the 3' end of the terminus of dsRNA. This distance is typically 21-25 nucleotides in length depending on the species and Dicer isoform. Dicer interacts with TAR RNA-binding protein (TRBP) which modulates the processing efficiency of some pre-miRNAs and adjusts the length of mature miRNA. A small 21-23 bp miRNA/miRNA* duplex generated by Dicer is subsequently loaded onto an effector complex, RISC[32]. Following duplex loading, RISC quickly unwinds the duplex and removes the passenger strand to generate a mature RISC. The guide strand is deter-

mined on the basis of the relative thermodynamic stability of the duplex strands. The strand with a relatively unstable 5' terminus is usually chosen as the guide strand[38].

1.3.2 Regulation of miRNA transcription and processing

Transcription of miRNA-encoding genes is regulated in a similar manner as the transcription of protein-coding genes. Regulatory transcription factors, such as p53, MYC, Zinc finger E-box-binding homeobox 1 (ZEB1) and ZEB2 and Myoblast determination protein 1 (MYOD1) positively or negatively regulate miRNA expression[38]. Epigenetic control, primarily DNA methylation and histone modification, are also major contributors to miRNA gene regulation[39].

Additionally, miRNAs can be regulated during their processing. Drosha and Dicer, two main processing enzymes, are assisted by a number of cofactors or accessory proteins which positively or negatively modulate their processivity[40].

Drosha and its cofactor DGCR8 are involved in a cross-regulatory loop. DGCR8 stabilizes Drosha through the interaction with its middle domain, whereas Drosha destabilizes *DGCR8* mRNA by cleaving hairpins present in the second exon[38]. This auto-regulation enables maintenance of homeostatic Microprocessor activity[32].

p68 and p72 are essential for Drosha-mediated processing of a subgroup of miRNAs. TGF-β-specific SMAD signal transducers interact with p68 and the stem of pri-miRNA to stimulate processing of miR-21 and miR-199a[41]. p53, enhances the post-transcriptional maturation of several growth-suppressive miRNAs, through association with p68[42].

Heterogeneous nuclear ribonucleoprotein A1 (HNRNPA1) and KH-Type Splicing Regulatory Protein (KSRP) facilitate Drosha-mediated processing through recognition of the terminal loop of pri-mir-18a and pri-let-7, respectively[43,44].

TRBP modulates the processing efficiency of some pre-miRNAs and stabilizes Dicer. Decrease in expression of TRBP leads to Dicer destabilization and pre-miRNA processing defects[38].

In many organisms, two or more different Ago proteins operate in the miRNA pathway; therefore control of miRNA function may occur at the level of Ago selection[39].

Ago proteins themselves can be tightly regulated. Mechanisms that have been implicated in Ago regulation include: polyubiquitylation and subsequent proteosomal degradation, hydroxylation and phosphorylation[45].

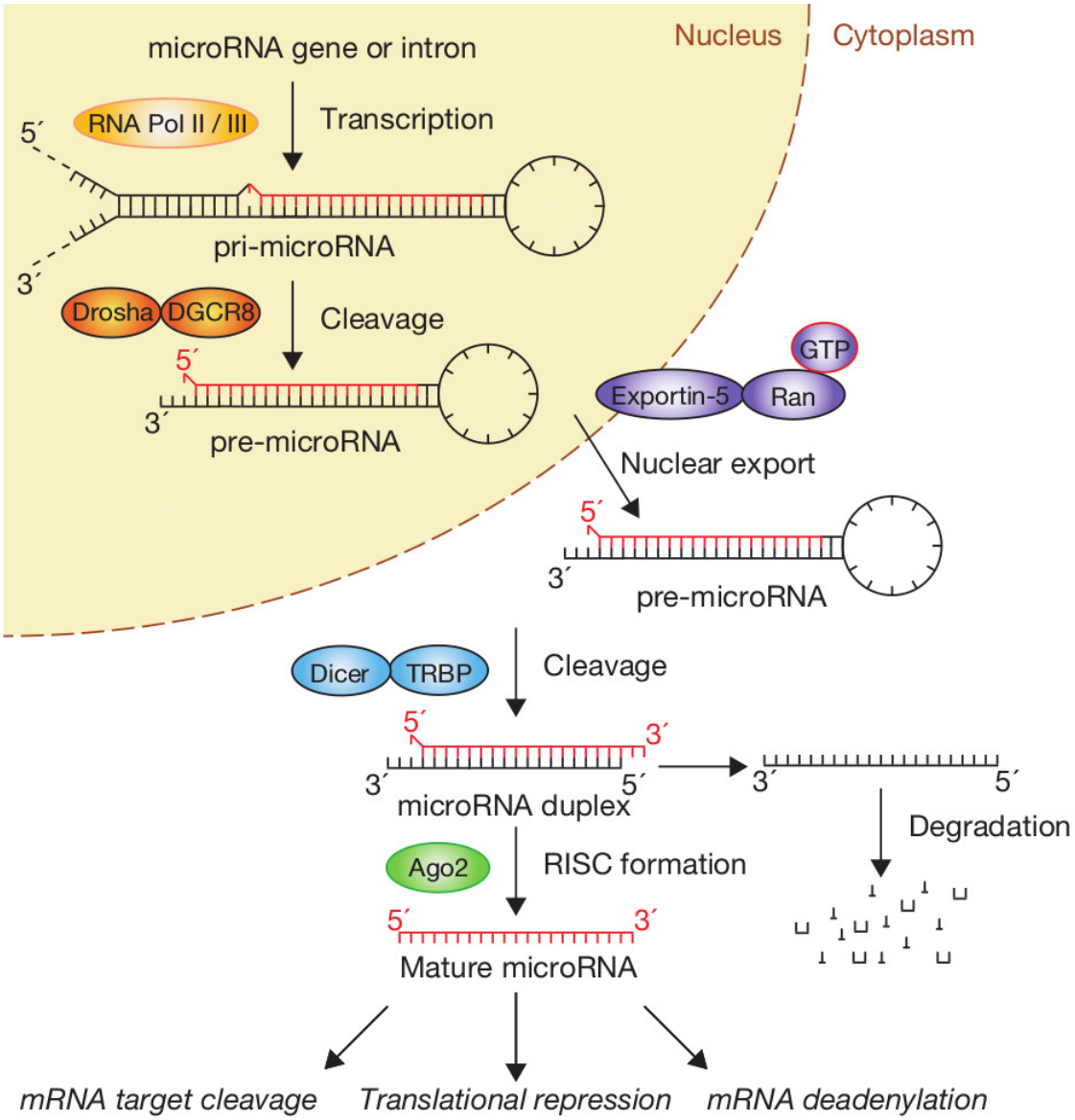

Figure 1.4: miRNA biogenesis. miRNAs are transcribed by RNA polymerase II or RNA polymerase III into pri-miRNA. The pri-miRNAs fold into hairpins, which act as substrates for Drosha, an RNase III family enzyme. Drosha cleaves pri-miRNA, producing a 70-nucleotide long pre-miRNA. The product of Drosha cleavage is exported to the cytoplasm, where it is processed by Dicer, another RNase III family enzyme. Dicer forms 21-23 bp miRNA/miRNA* duplex, one strand of which is then incorporated into the RISC. Reprinted from Winter et al., 2009[35].

Adenosine deaminases that act on RNA (ADARs) catalyse the conversion of adenosine to inosine in dsRNA segments. Many pri-miRNAs and pre-miRNAs are targeted by ADARs at different stages of their processing[46]. Selective editing can lead to inhibition of cleavage by Drosha or Dicer and thus to reduced expression of mature miRNA[47]. Editing within mature miRNA can have an important impact on the target specificity[48].

Another two major pathways that can influence miRNA stability and specificity are tailing and trimming[49].

Tailing involves addition of adenosine and uracil to the miRNA. While addition of adenosine or monouracil stabilizes miRNAs, addition of polyuracil marks them for degradation[50–52]. Trimming is a process in which the 5' or 3' end of miRNA is exonucleolytically cut. Trimming of 5' end has important functional consequences, as it can alter the miRNA seed region. Furthermore, it may affect the efficiency of miRNA loading into RISC since AGO loading typically selects for miRNAs with distinct 5' nucleotides[38]. miRNAs with an identical 5' end but different 3' end typically target the same mRNAs, but the efficiency of their targeting may differ or they may have different half-lives[46].

The expression of some miRNAs is regulated at the RNA stability level. Several nucleases have been considered as candidates involved in miRNA decay, but it is an open question how they achieve substrate specificity[53].

1.3.3 miRNAs and cancer

Global downregulation of miRNA expression is an emerging feature in cancer. Deregulation of certain miRNAs is seen to be tumour type specific[33]. Thus, alternations in miRNA expression profiles could be used to improve classification of cancer and predict tumour behaviour[54]. As miRNA expression increases with differentiation of the cell, apparent underexpression of miRNAs in cancer may be -in part- a result of shifting miRNA expression to a less-differentiated state allowing for tumour progression[55,56]. However, a clear-cut cellular and molecular mechanism has been described only for a small number of miRNAs. Possible explanations may include complex regulation of miRNA expression and function, variable influence on their targets, redundancy of miRNAs, heterogeneity of cells and technical limitations in small RNA profiling[57].

Additionally, studies describing roles that miRNAs have in regulation of fibrosis and HCC[23], typically have identified individual miRNA:target interactions. Whilst these

studies provide valuable insight in miRNA-directed regulation, such approaches ignore the complexity of miRNA signalling networks.

1.4 Serum response factor

Serum response factor (SRF) is an ubiquitously expressed transcription factor which regulates a wide range of biological processes. SRF controls and directs different developmental processes, functions of the heart and the cardiovascular system, liver regeneration, endothelial cell functions and vascularization, T and B cells activities and neuronal functions of the brain[58]. Homodimeric SRF recognizes and binds the consensus element CC(A/T)$_6$GG, called the CArG box[59].

SRF is recruited to the promoters of its targets genes together with different cofactors. Two principal families of SRF cofactors are ternary complex factor (TCF) and the myocardin family[60].

Members of the TCF, Elk1, Elk3 and Elk4, are part of the large ETS family of transcription factors. TCFs are activated through MAPK pathway. Activated TCFs stimulate cell proliferation, differentiation and migration[61]. These proteins recognize and bind Ets motif GGA(A/T) adjoining the CArG box.

Myocardin family members, myocardin, MRTF-A and MRTF-B, are controlled by Rho-family GTPases signalling pathway and monomeric actin. Monomeric actin binds MRTFs and retains them in cytoplasm. Polymerization of monomeric actin releases MRTFs which can interact with SRF once they are translocated to nucleus[58]. Although MRTFs contact DNA directly, for efficient target gene recognition and activation, they require SRF[59]. Genes regulated by MRTFs are involved in muscle-specific functions, actin dynamics, cell motility and serum response[62,63].

Ras/MAPK and Rho/actin signalling pathways are frequently activated in HCC. SRF, activated by both aforementioned cascades, is frequently elevated in human HCC. Therefore, it is not surprising that constitutive expression of SRF leads to HCC formation. To study HCC formation and development, Nordheim group has generated *SRF-VP16*^{iHep}mice[64]. These mice express in a hepatocyte-specific fashion constitutively active SRF due to fusion with the VP16 transcriptional activation domain of *Herpes simplex* virus. All *SRF-VP16*^{iHep}mice develop hyperproliferative liver nodules that progress to lethal HCC[64]. This animal model, which provides a perfect opportunity to study

formation and progression of HCC, was used in this study to describe the role of microRNAs in HCC development.

1.5 Aims

Considering the serious consequences of HCC occurrence and prognosis in humans, elucidation of the causes of HCC formation is necessary as it may suggest novel therapeutic strategies.

Owing to their capacity to modulate the expression of tens to hundreds of target genes, miRNAs are emerging as important factors in the control of cancer hallmarks[33]. Yet, due to the complexity of miRNA regulation, only a very small number of miRNAs have been identified as drivers of cancerogenesis.

The complexity of regulation of miRNAs and miRNA-directed targeting leads to intricate networks of miRNAs and their target and regulatory genes. The nodes of these networks, which can be either miRNAs or mRNAs, are generally connected to many other nodes in these regulatory networks. Hubs, nodes in the network with atypically high number of connections, are of special importance, as they represent sites of signalling convergence which can explain the network behaviour and serve as potential targets for therapy and prediction of clinical outcome[65].

In this study, the *SRF-VP16*^{iHep} mouse model was used as starting point to investigate the role of miRNAs in the regulation of HCC and associated microenvironment development.

The *SRF-VP16*^{iHep} mice were used to precisely characterize changes in miRNA profile at different tumour stages and to study how miRNA expression changes during tumor progression, from pre-malignant nodules to terminal HCCs.

Further analysis was performed to identify miRNA targets associated with carcinogenesis, thus providing mechanistic description of miRNA involvement in tumor and tumor microenvironment development. Special focus of the intended investigation on miRNA function in HCC was placed on the identification of a miRNA:mRNA network that influences microenvironment development in HCC with emphasis on miRNA hubs which regulate a considerable number of genes in the network.

Finally, studying the mechanism of miRNA dysregulation was performed with the aim of dissecting the complete role of miRNA involvement in HCC carcinogenesis.

Chapter 2

Material and Methods

2.1 Material

2.1.1 Animal models

SRF-VP16^{iHep} mouse model

The *SRF-VP16*^{iHep} mouse model, used in this study was generated by Dr. Stefan Ohrnberger in the group of Prof. Alfred Nordheim[64].

SRF-VP16^{iHep} mice carry and express a constitutively active form of SRF, SRF-VP16. The SRF-VP16 fusion protein is comprised of the first 412 residues of human SRF fused to the transcriptional activation domain of the *Herpes simplex* viral VP16 protein. The VP16 transcriptional activation domain enables SRF-directed target gene activation independently of upstream stimulation by Rho/actin and Ras/MAPK signalling pathways, thus rendering SRF-VP16 constitutively active[66].

To allow the conditional expression of SRF-VP16 upon cellular Cre recombinase activity, a floxed STOP cassette was introduced upstream of the SRF-VP16 coding sequence. The SRF-VP16 construct was integrated into the genomic Rosa26 locus[67] generating *Gt(ROSA)26-Sor*$^{tm1(SRF-VP16)Antu}$ mice.

To obtain the *SRF-VP16*^{iHep} mice stop-floxed SRF-VP16 mice were bred with *Srf-flex1* mice (floxed *Srf* exon 1) and *Alfp-CreER*^{T}2 animals which express tamoxifen-inducible hepatocyte-specific Cre recombinase. Therefore, this animal model shows conditional, Cre-mediated expression of SRF-VP16, combined with non-functional endogenous SRF. Tamoxifen treatment through activation of Cre recombinase efficiently induces SRF-VP16 expression in hepatocytes, which leads to hepatocyte hyperproliferation[64].

Although Cre activation is dependent on tamoxifen treatment, spontaneous, stochastic Cre activation was also observed, leading to mosaic SRF-VP16 expression in a small number of hepatocytes. Hyperproliferation of affected hepatocytes leads to development of pre-malignant nodules throughout the liver and their gradual progression to HCC[64].

The tissue samples of pre-malignant nodules and HCC used in this study were isolated from *SRF-VP16*^{iHep} mice, which developed HCC as a result of stochastic hepatocyte-specific activation of SRF-VP16. Liver tissue from litter siblings that did not develop HCC due to lack of either SRF-VP16 or Cre expression was used as controls. The tissue was provided by Dr. Abhishek Thavamani. The respective protocol was approved by Regierungspräsidium Tübingen (IM1/14 permit).

The carbon tetrachloride model

The murine carbon tetrachloride model (CCl_4) model serves as an experimental *in vivo* liver fibrosis model. CCl_4 (Sigma-Aldrich) was administered intraperitoneally to C57BL/6 mice (n=5) twice per week for six weeks at a concentration of 0.8 µl/(g body weight) diluted in mineral oil following standardised operation procedures[68]. Animals that received mineral oil alone served as controls. The respective protocol complied with the guidelines for animal care and was approved by the German Animal Care Committee and the Landesamt für Umwelt und Naturschutz (LANUV, Recklinghausen, Germany) under permit no. Az.: 84-02.04.2012.A092.

The liver tissue RNA samples of mice treated with the CCl_4 together with corresponding controls were provided by Prof. Ralf Weiskirchen.

2.1.2 Cell lines and primary cells

In this study the following cell lines were used: GRX, LX-2, NIH/3T3.

GRX cell line

The continuous murine hepatic stellate cell line GRX is an anchorage-dependent line displaying morphological characteristics of myofibroblasts (activated HSC)[69]. Upon retinol treatment it partially reverts its phenotype to that of quiescent stellate cells[70]. The cell line was obtained from the Rio de Janeiro Cell Bank (PABCAM, Federal University, Rio de Janeiro, Brazil).

GRX cells were kind gift of Prof. Ralf Weiskirchen.

NIH/3T3 cell line

Portal fibroblasts are additional minor source of ECM in the liver[23]. Therefore, as a standard fibroblast cell line mouse embryonic fibroblasts NIH/3T3 were used[71]. These cells when maintained under Tgf-β-rich conditions shows characteristics of myofibroblasts[72]. NIH/3T3 cells were obtained from the American Type culture collection (ATCC).

Primary HSC

Isolation of murine primary HSCs (pHSCs) from C57BL/6 mice was essentially performed as described previously[73]. Briefly, the liver of healthy mice weighing about 20-25 g was perfused with pronase-collagenase solutions through the portal vein. The cells of the digested livers were dispersed and filtered through a nylon mesh. pHSCs were further enriched and purified from the remaining cells by Nycodenz (Axis-Shield) density centrifugation. Cells were counted in a Neubauer chamber and the viability of cells determined by trypan blue exclusion method using a ready-to-use 0.4% trypan blue solution. Cells were finally seeded on uncoated plastic dishes in Dulbecco's modified Eagle's Medium (DMEM) supplemented with 10% fetal calf serum (FCS), 100 IU/mL penicillin, 10 µg/mL streptomycin and 4 mM glutamine. Permission to isolate pHSC from mice using the perfusion protocol was given by the LANUV under permit no.: Az.: 84.02.04.2015.A028.

Isolated pHSC were provided by Prof. Ralf Weiskirchen.

2.1.3 Vectors

To generate the pMSCV vector (vector without an gene insert in multiple cloning site), the pMSCV-Lin28A (Addgene.) vector was digested using *EcoRI* and *BglII* restriction enzymes under standard conditions (as described in section 2.2.2). Digested vector DNA was loaded onto 1 % agarose gel and the vector backbone was isolated from the gel using QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions.

Two single-stranded, complementary oligonucleotides were annealed to form a double-stranded DNA fragment. 3 µg of complementary oligonucleotides were mixed in 60 µl annealing buffer (100 mM NaCl, 50 mM HEPES, pH 7.4). Oligonucleotides were initially heated to 90°C for 4 min to dissolve secondary structures and gradually cooled to 37°C (0.03°C/s) to facilitate hybridization.

Annealed nucleotides were inserted into the vector backbone following the standard ligation protocol (section 2.2.2).

pSV Sport PPAR gamma 1 vector was modified by inserting the *Neomycine resistance gene* into the vector backbone.

pMSCV-mLin28A was a gift of George Daley (Addgene plasmid #26357)[74], pSV Sport PPAR gamma 1 was kindly provided by Bruce Spiegelman (Addgene plasmid #8886)[75] and pcDNA3-Egr1 by Eileen Adamson (Addgene plasmid #11729)[76].

pmirGLO vector was purchased from Promega.

Short description of all vectors used in this study is contained in Table 2.1. Vector maps are shown in Appendix D.

Vector	Vector description	Source
pMSCV-Lin28A	Expression of murine Lin28A	Gift of George Daley (Addgene plasmid#26357)
pmirGLO	pmirGLO Dual-Luciferase miRNA Target Expression Vector	Promega
pMSCV	empty pMSCV-Lin28A vector (without an insert)	Generated by C. Bitter
pmirGLO-Col1a1	Expression of *luciferase-Col1a1-3'UTR* gene construct	Generated by C. Bitter
pmirGLO-Pdgfa	Expression of *luciferase-Pdgfa1-3'UTR* gene construct	Generated by C. Bitter
pmirGLO-Adamts15	Expression of *luciferase-Adamts15-3'UTR* gene construct	Generated by C. Bitter
pmirGLO-Tgfbr1	Expression of *luciferase-Tgfbr1-3'UTR* gene construct	Generated by C. Bitter
pmirGLO-Col1a1-mut	Expression of *luciferase-Col1a1-3'UTR* gene construct with mutated let-7 and miR-29c target sites	Generated by C. Bitter
pmirGLO-Pdgfa-mut	Expression of *luciferase-Pdgfa1-3'UTR* gene construct with mutated miR-29c target site	Generated by C. Bitter
pmirGLO-Adamts15-mut	Expression of *luciferase-Adamts15-3'UTR* with mutated let-7 and miR-338 target site in pmirGLO	Generated by C. Bitter
pmirGLO-Tgfbr1-mut	Expression of *luciferase-Tgfbr1-3'UTR* gene construct with mutated let-7 target site	Generated by C. Bitter
pSV Sport PPAR gamma 1	Expression of mouse Pparg1	Gift of Bruce Spiegelman (Addgene plasmid #8886)
pcDNA3-Egr1	Expression of mouse Egr1	Gift of Eileen Adamson (Addgene plasmid #11729)

Table 2.1: List of used vectors, their description and vector's source

2.1.4 Primers

Primer name	Sequence (5'-3')	Target gene/miRNA
29c-m	AGCACCATTTGAAATCGGTTAAA	mmu-mir-29c-3p (mature)
29c-preF	TGACCGATTTCTCCTGGTGT	mmu-mir-29c (pre-mirna)
29c-preR	CCCCTACATCATAACCGATTTCA	mmu-mir-29c (pre-mirna)
29c-priF	CACATCAGTATGGCTCGTCTG	mmu-mir-29c (pri-mirna)
29c-priR	GTCTGAATCTCAAGGCAGGTG	mmu-mir-29c (pri-mirna)
338-m	TCCAGCATCAGTGATTTTGTTG	mmu-mir-338-3p (mature)
338-preF	GCCGTCCTCCCCAACAATA	mmu-mir-338 (pre-mirna)
338-preR	AGCTGCCCTCTTCAACAAAATC	mmu-mir-338 (pre-mirna)
338-priF	AAACTCACCTTGCAGACAGTT	mmu-mir-338 (pri-mirna)
338-priR	AGAGGCTGGGGTTGGTATG	mmu-mir-338 (pri-mirna)
335-m	TTTTTCATTATTGCTCCTGACCA	mmu-mir-335-3p (mature)
335-preF	GGGTCAAGAGCAATAACGAA	mmu-mir-335 (pre-mirna)
335-preR	TGGCTATAACCCATGAGAGG	mmu-mir-335 (pre-mirna)
335-priF	ACCCCATTCCCAAATTCATGC	mmu-mir-335 (pri-mirna)
335-priR	GCACCTATCTCCAAATGCTGT	mmu-mir-335 (pri-mirna)
30d-m	TGTAAACATCCCCGACTGGAA	mmu-mir-30d-5p (mature)
30d-preF	TGTGTCTGTAAACATCCCCGA	mmu-mir-30d (pre-mirna)
30d-preR	GAGCCAGTAGCAGCAAACATC	mmu-mir-30d (pre-mirna)
30d-priF	ACAGTTGTCATTCAGCTCCG	mmu-mir-30d (pri-mirna)
30d-priR	ACAGCAAGCTACAAAGGAACA	mmu-mir-30d (pri-mirna)
30d-m	TGTAAACATCCTTGACTGGAAGA	mmu-mir-30e-5p (mature)
30e-preF	ACTGTAAACATCCTTGACTGGAA	mmu-mir-30e (pre-mirna)
30e-preR	TGCCGCTGTAAACATCCGA	mmu-mir-30e (pre-mirna)
30e-priF	AAGTCTAGGAGAAGTGGGCAT	mmu-mir-30e (pri-mirna)
30e-priR	ACAGAGGGCTTGCTAATCCAA	mmu-mir-30e (pri-mirna)
let7a-m	TGAGGTAGTAGGTTGTATAGTTA	mmu-mir-let-7a-5p (mature)
let7a1-preF	GGGATGAGGTAGTAGGTTGT	mmu-mir-let-7a-1 (pre-mirna)

Continued on next page

Table 2.2 – *Continued from previous page*

Primer name	Sequence (5'-3')	Target gene/miRNA
let7a1-preR	ACAGTAGATTGTATAGTTATCTCCC	mmu-mir-let-7a-1 (pre-mirna)
let7a2-preF	TCCCAGGTTGAGGTAGTAGG	mmu-mir-let-7a-2 (pre-mirna)
let7a2-preR	GCAAGTCCCAAGGAAAGCTA	mmu-mir-let-7a-2 (pre-mirna)
let7a1-priF	TGGTACTCTGGGAACTAGGTTT	mmu-mir-let-7a-1 (pri-mirna)
let7a1-priR	CATCACAGGAAGCCGTTCTC	mmu-mir-let-7a-1 (pri-mirna)
let7a2-priF	TCTGTTTCATTCCCACCACG	mmu-mir-let-7a-2 (pri-mirna)
let7a2-priR	CTACCCATTCTCAAGCTCTGC	mmu-mir-let-7a-2 (pri-mirna)
let7c-m	TGAGGTAGTAGGTTGTATGGTT	mmu-mir-let-7c-5p (mature)
let7c1-preF	GGTTGAGGTAGTAGGTTGTATG	mmu-mir-let-7c-1 (pre-mirna)
let7c1-preR	CCAAGGAAAGCTAGAAGGTTG	mmu-mir-let-7c-1 (pre-mirna)
let7c2-preF	TTGGGGTGAGGTAGTAGGTT	mmu-mir-let-7c-2 (pre-mirna)
let7c2-preR	CCACTTCAGGAAAGACAGTAGAT	mmu-mir-let-7c-2 (pre-mirna)
let7c1-priF	AATTTGCTCGCTGCTAATGGA	mmu-mir-let-7c-1 (pri-mirna)
let7c1-priR	ACAGGACTACCCTAATCTGGTG	mmu-mir-let-7c-1 (pri-mirna)
let7c2-priF	CTCCCCTTACTCTCTCTGTGATT	mmu-mir-let-7c-2 (pri-mirna)
let7c2-priR	GATGTTGTGGCTTCTTGAGGT	mmu-mir-let-7c-2 (pri-mirna)
let7g-m	TGAGGTAGTAGTTTGTACAGTT	mmu-mir-let-7g-5p (mature)
let7g-preF	AGGCTGAGGTAGTAGTTTGT	mmu-mir-let-7g (pre-mirna)
let7g-preR	miScript universal primer	mmu-mir-let-7g (pre-mirna)
let7g-priF	GGTGTATTTCTTTTGTTGGGT	mmu-mir-let-7g (pri-mirna)
let7g-priR	ACACATTTGCCTCATCTTTAC	mmu-mir-let-7g (pri-mirna)
rnu6F	CTTCGGCAGCACATATACTAAA	rnu6 (U6 snoRNA)
rnu6R	ATGGAACGCTTCACGAATTT	rnu6 (U6 snoRNA)

Continued on next page

Table 2.2 – *Continued from previous page*

Primer name	Sequence (5'-3')	Target gene/miRNA
s33F	AGACATCTCCCACTCATGTTC	snord33
s33R	CAGCCTCAGATGGTAGTGC	snord33
Adamts14F	CGAGGCAGCATTTCCTCTAC	Adamts14
Adamts14R	TCTGAGTTGTCTGTGCGGAT	Adamts14
Adamts15F	CTGCGAGGGAGTAAGAGTGAA	Adamts15
Adamts15R	GCTGTGGTTGTAGCCATTGAA	Adamts15
Col1a1F	GATGCTAACGTGGTTCGTGA	Col1a1
Col1a1R	TGAGTAGGGAACACACAGGT	Col1a1
Col1a2-1f	TCAAAGGCGTGAAAGGACAC	Col1a2
Col1a2-1r	AGTTCCATTCTCTCCAGGGG	Col1a2
Col4a2-1f	GAGTGCGGTTCAAAGCGTC	Col4a2
Col4a2-1r	CATCCAACTTCTTCACACCCC	Col4a2
Col4a5-1f	CAGGTATCAAAGGGTCGGTG	Col4a5
Col4a5-1r	TCTTGACCTGGCTTGCCTTT	Col4a5
Col5a2-1f	GGCAAAGATGGAGAAGTTGGT	Col5a2
Col5a2-1r	CCCAGGCAGTCCAGTTATCC	Col5a2
Lamc1F	GCCTTCCTGACCGACTACAA	Lamc
Lamc1R	TCCCGAGTGCGCTTATAGAT	Lamc
pdgfaF	GATAGACTCCGTAGGGGCTG	Pdgfa
pdgfaR	ACTTCTCTTCCTGCGAATGG	Pdgfa
pdgfbF	GATCTCTCGGAACCTCATCG	Pdgfb
pdgfbR	GGGCTTCTTTCGCACAATCT	Pdgfb
tgfbr1F	CATTTCAGAGGGCACCACCT	Tgfbr1
tgfbr1R	CAAACTTCTCCAAACCGACCT	Tgfbr1
tpm1F	TCAAGGTTCTCTCTGACAAGC	Tpm1
tpm1R	TTGGTTACTGATCTCTCTGCAA	Tpm1
Gapdh-f	TGGATCTGACGTGCCGC	Gapdh
Gapdh-r	ATGCCTGCTTCACCACCTTC	Gapdh
Gusb-f	GATGCTGTTCCCGAAGGAGAG	Gus

Continued on next page

Table 2.2 – *Continued from previous page*

Primer name	Sequence (5'-3')	Target gene/miRNA
Gusb-r	ATTGCTTCCCGTTCATACCACA	Gus
TBPF	TTCATGGTGTGTGAAGATAACCC	Tbp
TBPR	AGAGAGACTGTTGGTGTTCTGA	Tbp
Lin28a-f	GGTGGTGTGTTCTGTATTGGG	Lin28a
Lin28a-r	TAGGCAGGCTTTCCCTGAG	Lin28a
adamts15-utrF	CGGCTAGCTCCTTTATTCTCCACCCGCT	luciferase assay: Adamts15-UTR
adamts15-utrR	GCGTCGACAGGAAGAAGATACCCTACCC	luciferase assay: Adamts15-UTR
col1a1-utrF	CGAGCTCACTGGAAACATCGGACATGC	luciferase assay: Col1a1-UTR
col1a1-utrR	GCGTCGACTCTAGGGAGCATCTCAGCTTC	luciferase assay: Col1a1-UTR
pdgfa-utr2F	CGAGCTCATGTGCGTGGCGTGTGACATT	luciferase assay: Pdgfa-UTR
pdgfa-utr2R	GCGTCGACCGCTCTCTGTGACAAGGAAG	luciferase assay: Pdgfa-UTR
tgfbr1-utr1F	CGAGCTCCTGAAACACCGTGGGAACTCT	luciferase assay: Tgfbr1-UTR
tgfbr1-utr2R	GCGTCGACTACAACGGAGCAACTGTGCC	luciferase assay: Tgfbr1-UTR
NeoF	CGAGCTCTGCTCCCAGAAGCC	Neomycine gene
NeoR	CAGGTCGACGGATCCCCTCAGAA	Neomycine gene
Col1a1-let7F	AACGTTCTTGTCTTTGTGTGTC	luciferase assay; mutagenesis (Col1a1:let-7 site)
Col1a1-let7R	TTTGTCTTTCAGCAACACAATTG	luciferase assay; mutagenesis (Col1a1:let-7 site)
Col1a1-29c-1F	ATTCATCTCTGTGAAGGGGTG	luciferase assay; mutagenesis (Col1a1:miR-29c site 1)

Continued on next page

Table 2.2 – *Continued from previous page*

Primer name	Sequence (5'-3')	Target gene/miRNA
Col1a1-29c-1R	CGAATTTGATACCAAACTGGGC	luciferase assay; mutagenesis (Col1a1:miR-29c site 1)
Col1a1-29c-2F	ATTTAGAAGCTGAGATGCTCC	luciferase assay; mutagenesis (Col1a1:miR-29c site 2)
Col1a1-29c-2R	CGAAAGAAATTCCTTCCCACCC	luciferase assay; mutagenesis (Col1a1:miR-29c site 2)
TGFB-let7F	AACACTGAGAGAACAGAGGGC	luciferase assay; mutagenesis (Tgfbr1:let-7 site)
TGFB-let7R	TTTGAACAACCAGCCTCCTAAAC	luciferase assay; mutagenesis (Tgfbr1:let-7 site)
PDGF-29cF	CATTCATTGCCAATGTGCGTG	luciferase assay; mutagenesis (Pdgfa:miR-29c site)
PDGF-29cR	TAAGTACATAGTATGTTCAGGAATGTC	luciferase assay; mutagenesis (Pdgfa:miR-29c site)
Adamts-29cF	GAAGTAACCTCCAAAACTCACCCGTC	luciferase assay; mutagenesis (Adamts15:miR-29c site)
Adamts-29cR	TGTCCTAGAGAGCATGATC	luciferase assay; mutagenesis (Adamts15:miR-29c site)
Adamts-338F	AAAAGAGCCATGGCCTGCAGC	luciferase assay; mutagenesis (Adamts15:miR-338 site)
Adamts-338R	TTAGGCCTGTGACCTCCATTATGGG	luciferase assay; mutagenesis (Adamts15:miR-338 site)
Adamts-let7-1F	AAATTGCCCTGCTTCTCGGGCC	luciferase assay; mutagenesis (Adamts15:let-7 site 1)
Adamts-let7-1R	TTAGACTGTGCATTAGTTAACTTCCCCCTG	luciferase assay; mutagenesis (Adamts15:let-7 site 1)
Adamts-let7-2F	AAATGATCATGCTCTCTAGGAC	luciferase assay; mutagenesis (Adamts15:let-7 site 2)
Adamts-let7-2R	TTAGCAAACAATGAAGCCCGTTTAC	luciferase assay; mutagenesis (Adamts15:let-7 site 2)

Continued on next page

Table 2.2 – *Continued from previous page*

Primer name	Sequence (5'-3')	Target gene/miRNA
30e-1F-2	CCAAAAGTCTCCGCTCTCACTATC	ChIP; EGR1 and PPARG
30e-1R-2	GGGGAAGCGTGCGTCA	ChIP; EGR1 and PPARG
Let7a-2F-2	TTCACATTCCTCACCTCGCC	ChIP; PPARG
Let7a-2R-2	CAAACCTGACGGCGGAAG	ChIP; PPARG
29c-1F	ACCTACCTCCTCCTCACCTTC	ChIP; PPARG
29c-1R	TCTCCTGGCTCCACTGAATC	ChIP; PPARG
338-1F	CGGTCTGTAACCTCACCTGT	ChIP; PPARG
338-1R	GAGAGTACGGAGGTCGAAGG	ChIP; PPARG
335-1F	GCAAACCAGGCTCCGAAGA	ChIP; EGR1
335-1R	AGCTAGGGATGCTGCGAAC	ChIP; EGR1
30d-1F	TCTGAACGTCCTCCTTCACC	ChIP; PPARG
30d-1R	GGTAGTGCGAAAACTCTCCC	ChIP; PPARG
let7a-1F	GTCCGCCAATAGGAATGCG	ChIP; EGR1
let7a-1R	GCGAGAACCCGGATGGAA	ChIP; EGR1
Let7a-2F	GGGATGGCCTGGGCG	ChIP; PPARG
Let7a-2R	GGGCGAGGTGAGGAATGT	ChIP; PPARG
let7c-1F	TGTGGTATCAGGAGACTGCC	ChIP; PPARG
let7c-1R	TCTGGCACTGTATGGTCCTT	ChIP; PPARG
Let7c-2F	CCGTGCCCCAGTCCTAATA	ChIP; EGR1
Let7c-2R	GATTCGCGGAGGCATGTT	ChIP; EGR1
let7g-1F	ACACCCCGGAACAGATTGTA	ChIP; EGR1 and PPARG
let7g-1R	ACGACGAAATGGAACACGAG	ChIP; EGR1 and PPARG
mir-29c-1F	TTTGTAATATAAGGAGTAGATTGG	methlyation assay; miR-29c
mir-29c-1R	TACCAACTCTTATACTCACTCC	methlyation assay; miR-29c
mir-29c-2F	GGGTAAAGGGTGAAAGGTTAAG	methlyation assay; miR-29c
mir-29c-2R	CCTTATATCTAAACAAAAATTTAAAAC	methlyation assay; miR-29c
mir-338-1F	GTTTTYGGATTGTATTAGTTAGAG	methlyation assay; miR-338
mir-338-1R	ATACAAAATCTAACCCCAAAACC	methlyation assay; miR-338

Continued on next page

Table 2.2 – *Continued from previous page*

Primer name	Sequence (5'-3')	Target gene/miRNA
mir-338-2F	ATGAGTTTTTGGATTTTTTATTTAG	methlyation assay; miR-338
mir-338-2R	AAATCTACTACTAAATAAATAAAAATT	methlyation assay; miR-338
mir-338-3F	GATTTAGGAATTTATGGTTATAGG	methlyation assay; miR-338
mir-338-3R	AACACTCCCTCTCCCCCAATC	methlyation assay; miR-338
mir-335-1F	ATAGGGTTTGGTTTTAAGGTTTG	methlyation assay; miR-335
mir-335-1R	CTTCAACCCCAACTTTACCTTC	methlyation assay; miR-335
mir-335-2F	GTTTTGTTTTAGAAAGTAATTTTTAG	methlyation assay; miR-335
mir-335-2R	ACRACATAACCAAAACCAACAAC	methlyation assay; miR-335
mir-335-3F	GTAATTTTGTAATTATGAGTTTTTGG	methlyation assay; miR-335
mir-335-3R	CCAAAAAAAATCTACTACTAAATAAAT	methlyation assay; miR-335
mir-335-4F	GGATTTAGGAATTTATGGTTATAG	methlyation assay; miR-335
mir-335-4R	AATCTTAAAACCCTCCTACCAAC	methlyation assay; miR-335
mir-30d-1F	YGGGGGTTTATTTGTTTTAGGG	methlyation assay; miR-30d
mir-30d-1R	CTCTCCTCAAACACACTAAAA	methlyation assay; miR-30d
mir-30d-2F	TAGATATGTTTAGTGATGGAGGG	methlyation assay; miR-30d
mir-30d-2R	TTTAACCTTTCCCTCCAATATATC	methlyation assay; miR-30d
mir-30d-3F	TAGTAGAAAGAGGTTGGATATTG	methlyation assay; miR-30d
mir-30d-3R	CACTATAATCATTTAATATCACTCC	methlyation assay; miR-30d
mir-30d-4F	AAGAGGAGAATAAAAATTTGTAGG	methlyation assay; miR-30d
mir-30d-4R	AAACTTCTTTCCCACTTAACCAC	methlyation assay; miR-30e

Continued on next page

Table 2.2 – *Continued from previous page*

Primer name	Sequence (5'-3')	Target gene/miRNA
mir-30e-1F	GTTTTTYGTTAATGGGAGAAAG	methlyation assay; miR-30e
mir-30e-1R	AACRAAAAACTCCCTTCCCTC	methlyation assay; miR-30e
mir-30e-2F	TTATTTAAGTTTTTATATATTTATAGG	methlyation assay; miR-30e
mir-30e-2R	CTCTCCCTACCATTCATTTCC	methlyation assay; miR-30e
mir-30e-3F	AGGAAATGAATGGTAGGGAGAG	methlyation assay; miR-30e
mir-30e-3R	GTAAGTGAAGAAATAAAATTTTTAAAT	methlyation assay; miR-30e
mir-30e-4F	ATTTGTTAGTTTAGTTAAATAAGTAAG	methlyation assay; miR-30e
mir-30e-4R	CATATATTACCACACCAAACCATAC	methlyation assay; miR-30e
mir-30e-5F	GGAAGTTAATTATTTTGGGGAGG	methlyation assay; miR-30e
mir-30e-5R	AACATAATAATTTATATCTATAATCCC	methlyation assay; miR-30e
let7a1-1F	GAGGTGAGGAATGTGAAGAATG	methlyation assay; let-7a
let7a1-1R	CATTTTTCCCTCCTCCAAACTC	methlyation assay; let-7a
let7a1-2F	GGYGGTTTGAGGGAAGGTTA	methlyation assay; let-7a
let7a1-2R	TACACCATAAAAAACCTAACTCC	methlyation assay; let-7a
let7a1-3F	TGTTATTAATTTTTTATATAAAGTTGG	methlyation assay; let-7a
let7a1-3R	AATTACAATTACTTATTTATCTAAAC	methlyation assay; let-7a
let7a2-1F	GGGTTAATTTTTGTTTTAGAGGG	methlyation assay; let-7a
let7a2-1R	CTACTACTATATATCCAATTTAAAC	methlyation assay; let-7a
let7a2-2F	ATAGAGGTTGTTGTTTATAGAAGG	methlyation assay; let-7a
let7a2-2R	ATCTCACAATTAACTTAATAAAAACC	methlyation assay; let-7a
let7c1-1F	AGTTGATTAAAGTTAATTTAGTTTTG	methlyation assay; let-7c
let7c1-1R	AACTATATTAATTACAATTTCCTCC	methlyation assay; let-7c
let7c1-2F	GGAAATTGTAATTAATATAGTTTTG	methlyation assay; let-7c
let7c1-2R	CTACTTATTACAACTTAAATAACC	methlyation assay; let-7c
let7c1-3F	TATATTTTATATATTTATGTTGTGTG	methlyation assay; let-7c

Continued on next page

Table 2.2 – *Continued from previous page*

Primer name	Sequence (5'-3')	Target gene/miRNA
let7c1-3R	CTAACTATAATTATCCAACAACC	methlyation assay; let-7c
let7c2-1F	TGAAAATTATAAAGTTAATATTTTGG	methlyation assay; let-7c
let7c2-1R	ATTCTCCCACAAAACTACTCTC	methlyation assay; let-7c
let7c2-2F	TTTTAGGGTTTTATTAGATTATAAG	methlyation assay; let-7c
let7c2-2R	CTCTTCAATATTATACCAAAATAAC	methlyation assay; let-7c
let7c2-3F	GTTTATAGTATTTAGATTTTAATAGG	methlyation assay; let-7c
let7c2-3R	CAACCACTTAAACCTAACTAATTC	methlyation assay; let-7c
let7c2-4F	TTTGAGGTTTAGGTTGGTTTGTG	methlyation assay; let-7c
let7c2-4R	ATTACCCAAAACTAATACCCACC	methlyation assay; let-7c
let7g-1F	TGAATGAAGTTAAATTGAATGATG	methlyation assay; let-7g
let7g-1R	AAATAACAAAAACAAATCTAATACC	methlyation assay; let-7g
let7g-2F	TTTGGAAAGTTGTTATTTATTTAGG	methlyation assay; let-7g
let7g-2R	ATTAATTAAATAATATACAACATACAC	methlyation assay; let-7g
let7g-3F	TTTTGAAGGGAGAAGTAAAATTAG	methlyation assay; let-7g
let7g-3R	CAAATATATATTATACATCACTCAAC	methlyation assay; let-7g
let7g-4F	GTYGGTTGGTTTTGAATTGATTT	methlyation assay; let-7g
let7g-4R	CTCACTTACCCTCCTAACAATC	methlyation assay; let-7g

Table 2.2: List of used primers, their sequence and targeted gene/miRNA

Amplicon	Amplicon mm10 coordinate	Amplicon size (bp)
mir-29c-1	chr1:195,016,936-195,017,402	508
mir-29c-2	chr1:195,017,459-195,017,770	353
mir-338-1	chr11:120,046,652-120,047,172	562
mir-338-2	chr11:120,047,847-120,048,193	388
mir-338-3	chr11:120,048,179-120,048,651	514
mir-335-1	chr11:120,046,654-120,046,909	297
mir-335-2	chr11:120,046,917-120,047,254	379
mir-335-3	chr11:120,047,841-120,048,207	408
mir-335-4	chr11:120,048,178-120,048,634	498
mir-30d-1	chr15:68,363,152-68,363,582	506
mir-30d-2	chr15:68,363,152-68,363,582	472
mir-30d-3	chr15:68,363,639-68,364,114	517
mir-30d-4	chr15:68,364,175-68,364,503	370
mir-30e-1	chr4:120,825,296-120,825,451	197
mir-30e-2	chr4:120,825,969-120,826,214	526
mir-30e-3	chr4:120,825,969-120,826,214	287
mir-30e-4	chr4:120,826,575-120,826,878	345
mir-30e-5	chr4:120,826,909-120,827,085	218
let7a1-1	chr13:48,546,030-48,546,309	321
let7a1-2	chr13:48,546,493-48,546,984	533
let7a1-3	chr13:48,547,511-48,547,954	485
let7a2-1	chr9:41,474,025-41,474,517	534
let7a2-2	chr9:41,474,683-41,474,906	265
let7c1-1	chr16:77,328,083-77,328,571	530
let7c1-2	chr16:77,328,550-77,329,079	571
let7c1-3	chr16:77,329,098-77,329,611	555
let7c2-1	chr15:85,668,406-85,668,755	391
let7c2-2	chr15:85,668,905-85,669,207	344
let7c2-3	chr15:85,669,446-85,669,851	447
let7c2-4	chr15:85,669,926-85,670,222	338
let7g-1	chr9:106,169,614-106,169,942	370
let7g-2	chr9:106,170,049-106,170,549	542
let7g-3	chr9:106,170,533-106,170,950	472
let7g-4	chr9:106,170,977-106,171,384	449

Table 2.3: List of amplicons generated in the DNA methylation analysis, their sizes and genome coordinates

2.1.5 Reagents and commercial assay (kits)

Chemicals and enzymes	Source	Experiment
Phosphate buffered saline (PBS)	Sigma-Aldrich	cultivation of cell lines
Trypsin-EDTA, 0.05 %	Gibco	cultivation of cell lines
Dulbecco's Modified Eagle Medium (DMEM), high glucose	Gibco	cultivation of cell lines
Penicillin-streptomycin (P/S), concentration 10 000 U/ml	Gibco	cultivation of cell lines
Fetal calf serum (FCS)	"Bio&SELL"	cultivation of cell lines
Trypan Blue, 0.4 %	Thermo Scientific	cell counting
Dimethyl sulfoxide (DMSO)	AppliChem	cell freezing, cloning
Salmon sperm DNA, 10 µg/µl	Invitrogen	electroporation
TransIT-LT1	Mirus Bio	transfection
RNAiMAX	Thermo Scientific	transfection
Opti-MEM I Reduced-Serum Medium	Thermo Scientific	transfection
G418	Invivogen	stable cell selection
Hydrochloric acid (HCl)	AppliChem	chemical treatment
Formaldehyde, 16 %	Thermo Scientific	ChIP
Glycine	AppliChem	ChIP
Protease inhibitor	Roche	ChIP
Sodium chloride (NaCl)	Sigma-Aldrich	ChIP, cloning
Ethylenediaminetetraacetic acid (EDTA)	AppliChem	ChIP
Tris(hydroxymethyl)aminomethane (Tris)	Sigma-Aldrich	ChIP
NP40	Fluka Biochemika	ChIP

Continued on next page

Table 2.4 – *Continued from previous page*

Chemicals and enzymes	Source	Experiment
Sodium dodecyl sulfate (SDS)	AppliChem	ChIP
Triton x-100	AppliChem	ChIP
Anti-PPAR/$gamma$ (PP-A3409A)	Perseus proteomics	ChIP
Anti-Egr1 (#4153)	Cell signaling	ChIP, immunostaining
IgG	Milipore	ChIP
Protein A-coupled Dynabeads	Thermo Scientific	ChIP
Polyvinylpyrrolidone (PVP)	Sigma-Aldrich	ChIP
Albumin (BSA)	AppliChem	ChIP
TWEEN-20	AppliChem	ChIP
Phenol:chloroform	AppliChem	ChIP
Chloroform	AppliChem	ChIP
Sodium acetate	Thermo Scientific	ChIP
Glycogen, RNase-free, 20 µg/µl	Thermo Scientific	ChIP
Phusion HF Buffer, 5X	NEB	cloning
Magnesium chloride (MgCl2)	NEB	cloning
Deoxynucleotide triphosphates (dNTPs), 10 mM	NEB	cloning
Phusion high-fidelity polymerase	NEB	cloning
Agarose LE, DNA grade	Genaxxon	cloning, PCR
XbaI	NEB, Fermentas	cloning
SacI	NEB, Fermentas	cloning
SalI	Fermentas	cloning
EcoRI	NEB	cloning
BglII	NEB	cloning

Continued on next page

Table 2.4 – *Continued from previous page*

Chemicals and enzymes	Source	Experiment
Tango Buffer, 10X	Fermentas	cloning
NEBuffer3.1, 10X	NEB	cloning
CutSmart Buffer, 10X	NEB	cloning
Calf intestine alkaline phosphatase (CIAP)	Fermentas	cloning
T4 DNA Ligase	NEB	cloning
T4 DNA Ligase Buffer, 10X	NEB	cloning
Tryptone	AppliChem	cloning
Yeast extract	AppliChem	cloning
Agar bacteriology grade	AppliChem	cloning
Glycerol	AppliChem	cloning
Polyethylene glycol (PEG) 8000	AppliChem	cloning
Magnesium sulfate ($MgSO_4$)	Sigma-Aldrich	cloning
Magnesium chloride ($MgCl_2$)	Merck	cloning
Potassium chloride (KCl)	Merck	cloning
Ampicillin, 100 mg/ml (Amp)	AppliChem	cloning
TRIzol Reagent	Ambion	RNA isolation
RNase-free H_2O	Thermo Scientific	RNA isolation

Continued on next page

Table 2.4 – *Continued from previous page*

Chemicals and enzymes	Source	Experiment
1-Butanol	Sigma-Aldrich	RNA isolation
Diethylether	Sigma-Aldrich	RNA isolation
Ethidium bromide (EtBr)	AppliChem	RNA isolation
FastStart Universal SYBR Green Master Mix	Roche	qPCR
Rotihistol	Carl Roth	immunostaining
Anti-ACTA2 (ab7817)	Abcam	immunostaining
4',6-diamidino-2-phenylindole (DAPI)	Thermo Scientific	immunostaining
Picro-Sirius Red	Sigma-Aldrich	immunostaining

Table 2.4: List of used chemicals, their source and experiment in which they were used

Kits	Manufacturer
DNeasy Blood & Tissue Kit	Qiagen
QIAquick PCR Purification Kit	Qiagen
QIAquick Gel Extraction Kit	Qiagen
EndoFree Plasmid Maxi Ki	Qiagen
mirVana Isolation Kit	Ambion
RNAse-free DNase Treatment and Removal Reagents Kit	Thermo Scientific
dsDNA BR Assay Kit	Thermo Scientific
RNA BR Assay Kit	Thermo Scientific
Q5 Site-Directed Mutagenesis kit	NEB
Dual-Luciferase Reporter Assay	Promega
miScript SYBR Green PCR Kit	Qiagen
miScript II RT Kit	Qiagen
TruSeq RNA sample preparation kit	Illumina
TruSeq Small RNA Library Prep Kits v2	Illumina
MaxFlour 10/31 Immunofluorescence Detection Kit	Dianova

Table 2.5: List of used kits and their source

2.1.6 Software

Software	Source
Olympus Fluoview (FV1000)	Olympus
ImageJ v1.51n	NIH
FASTQc v0.11,4.	Babraham Bioinformatics
STAR v2.5.2b	Dobin et al., 2013
R v3.5.1	N/A
RStudio v1.1.463	N/A
DESeq2 v1.22.2	Love et al., 2014
ggplot2 v3.1.0	N/A
circlize v0.4.5	Gu Z, 2014
LinReg v2018.0	Ruijter et al., 2013
FIMO v4.12.0	Grant et al., 2011
HOMER v4.10	Heinz S et al., 2010
MARS Data Analysis Software v1.20	BMG LABTECH
OPTIMA software v2.20	BMG LABTECH
QuantStudio Real-Time PCR software v1.0	Thermo Scientific
EpiTYPER v1.3	Agena CX

Table 2.6: List of used software and its source

2.2 Methods

2.2.1 Cell culture techniques

Cultivation of cell lines

All cell lines were maintained under standard growing conditions at 37°C in a humidified atmosphere with 5% CO_2. Routinely, the cells were grown in DMEM culture medium supplemented with 10% Fetal calf serum (FCS), 100 IU/mL penicillin, 10 µg/mL streptomycin and 4 mM glutamine.

All cell handling was done under sterile conditions.

To passage the cells, cells were washed in PBS to remove all traces of FCS which inhibits trypsin used in the sequential step. After washing, cells were detached in

trypsin for 2-5 min and collected in culture medium. To remove the trypsin residue completely, cells were centrifuged for 5 min at 800 rpm and room temperature. The cell pellet was re-suspended in complete culture medium and seeded in appropriate density in the new culture dish or harvested for the purpose of the experiment.

Cell counting

To determine the living cell count, a Neubauer counting chamber (depth of 0.1 mm) was used. A small aliquot of cell suspension was mixed with 0.4% trypan blue in ratio 1:2. Trypan blue selectively stains the dead cells and thus allows to determine the cell viability. The cell concentration was determined using the following formula:

$$\text{cell concentration}[\frac{\text{cells}}{\text{ml}}] = \frac{\text{number of counted cells}}{\text{number of counted squares}} * 10^4 * \text{dilution factor} \quad (2.1)$$

Freezing and thawing of cells

For freezing, cells were detached using trypsin, collected by centrifugation and re-suspended in freezing medium (50 % culture medium, 40 % FCS, 10 % dimethyl sulfoxide (DMSO)). The cell suspension was transferred to cryotubes and frozen in isopropanol freezing containers at -80 °C to ensure slow freezing (1°C/min). For long term storage, cells were transferred into liquid nitrogen container.

To thaw the cells, cryotubes containing the cells were immersed into a 37°C water bath for short period of time. The thawed cell suspension was added drop-wise to culture medium in a culture dish. Cells were allowed to attach overnight. The following morning, medium was changed to remove the cell-toxic DMSO.

Cell transfection

The cells were transfected using lipofection or electroporation procedure.

Electroporation introduces DNA into cells through usage of a high-voltage electric field which leads to the formation of pores in the cell membrane large enough to allow DNA to enter the cells[77].

Electroporation was based on the protocol by Sambrook and Russell[78] with all parameters optimised for the cell lines used in this study.

To perform the electroporation, 10^6 cells were used per experiment. The electroporation pulse generator was set to an impulse of 320 V for 15 ms. To transfect the cells, 20 μg of vector DNA was used together with 100 μg of salmon sperm carrier DNA. Following electroporation, cells were incubated for 24 h.

To transfect vector DNA and miRNA mimics, lipofection reagent TransIT-LT1 and RNAiMAX reagents were used, respectively.

The lipofection procedure uses a lipopolyplex transfection reagent, which forms lipid:DNA (or lipid:RNA) complex with vector DNA/RNA. The complex fuses with the cell membrane to release DNA/RNA into the cell[79].

24h prior to lipofection, cells were harvested and seeded in density of 200 000 cells per well of 6-well plate. The following day, 1 μg vector was added to 250 μl of Opti-MEM I Reduced-Serum Medium together with 2.5 μl of TransIT-LT1 Reagent per well. The solution was incubated for 15 min at room temperature. The Reagent:DNA complexes were added in a drop-wise manner to the cells and cells were incubated for 24 h.

miRNA mimics were used to modulate miRNA expression. miRNA mimics are double-stranded, chemically modified RNAs intended to mimic the function of endogenous miRNA[80]. To transfect the mimics, cells were seeded in density of 200 000 cells per well of 6-well plate. After 24h period, 150 μl of OPTI-MEM medium was mixed with 9 μl of RNAiMAX reagent per well. Appropriate amount of miRNA mimic was mixed with equal volume of OPTI-MEM medium. Diluted miRNA mimic was added to diluted RNAiMAX reagent in ratio 1:1 and solution was incubated for 5 min. The Reagent:RNA complexes were added in a drop-wise manner to the cells and cells were incubated for 24 h.

In luciferase assay, the final concentration of all miRNA mimics was 50 μM. In transfection experiments coupled with quantitative PCR (qPCR) the final concentrations of miR-29c and let-7g mimics were 100 μM while the miR-338 and let-7c mimics were used at a final concentration of 50 μM.

If miRNA mimics were to be transfected together with the DNA vector (e.g. as in luciferase assay), DNA vector was first transfected using the TransIT-LT1 Transfection protocol. At least 4h post-transfection, medium was changed and RNAiMAx protocol listed above was performed.

To generate stable cell lines, vector DNA was transfected using the electroporation procedure. 24 h post-transfection, the cell culture medium was exchanged with a fresh medium supplemented with selection antibiotic (NIH/3T3 cells - 600 μg/ml G418 and GRX cells - 700 μg/ml G418). Selection medium was changed every 2 to 3 days. Cells were expanded over 2 to 3 weeks to ensure the survival of only stably transfected cells.

Activation of pHSC

pHSC activation, evident by loss of retinoid droplets and increased ECM production, occurs when cells are plated on standard tissue plastic[14].

To ensure the full activation of pHSC, cells were maintained in DMEM culture medium supplemented with 10% FCS, 100 IU/mL penicillin, 10 μg/mL streptomycin and 4 mM glutamine for seven days. During this period cells were passaged once (after four days of cell culture) using the standard procedure for cell culture maintenance.

2.2.2 Molecular cloning techniques

Preparation of DNA insert

Genomic DNA (gDNA) was isolated from NIH-3T3 cells using the DNeasy Blood & Tissue Kit according to the manufacturer's instructions. Isolated gDNA served as a template in amplification of 3'-UTRs of selected miRNA target genes.

DNA fragment was amplified by polymerase chain reaction (PCR). To ensure that the sequence of the cloning insert is correct PCR was performed using Phusion polymerase with the corresponding reagents listed in Table 2.7. The approach was prepared on ice.

The PCR reaction was incubated in the thermocycler using the settings listed in Table 2.8.

Purification of PCR amplicon was performed using the QIAquick PCR Purification Kit according to the manufacturer's instructions. Alternatively, if PCR resulted in amplification of unspecific product, DNA fragments were purified from agarose gels using QIAquick Gel Extraction Kit. Isolated PCR product was inserted in pmirGLO vector using the procedure described in the following paragraphs.

Component	Reagent amount
5X Phusion HF Buffer	10 µl
50 mM $MgCl_2$	0.5 µl
10 mM dNTPs	1 µl
2 µM forward and reverse primer	5 µl
Phusion Polymerase (2 U/µl)	0.5 µl
Template DNA	100 ng
H_2O	to 50 µl

Table 2.7: PCR approach using Phusion DNA polymerase-components

	Step	Temperature	Time
	Initial Denaturation	98 °C	30 s
40 cycles	Denaturation	98 °C	10 s
	Primer annealing	45°C-72°C	30 s
	Elongation	72°C	30 s
	Final elongation	98 °C	10 min
	Hold	4 °C	

Table 2.8: Thermocycler conditions for Phusion DNA polymerase

Restriction digestion

To facilitate the cloning process restriction endonucleases were used to cleave double-stranded DNA at specific restriction sites, generated by PCR, thus creating nucleotide overhangs and allowing for more efficient ligation of DNA fragments and vector backbone. To create the overhangs, the DNA of interest, vector and insert alike, were incubated with the required endonuclease in a standard reaction described in Table 2.9. The approach was prepared on ice.

Digestion of DNA by restriction endonucleases was generally performed for 1 h at the temperature recommended by the manufacturer. If the generation of two different nucleotide overhangs was required, a double digest with two endonucleases was performed. Used endonucleases are listed in Chemical and enzymes table. Primers used to generate the restriction sites and amplify PCR product are listed in Table 2.2.

Component	Reagent amount
Endonuclease	10 U
DNA	1 µg
10x Buffer	5 µl
H_2O	to 50 µl

Table 2.9: Standard reaction for restriction digestion

Ligation

To avoid vector self-ligation in subsequent reactions, calf intestine alkaline phosphatase (CIAP) was added to the reaction mix in a concentration of 1 U/µg DNA.

The endonucleases were heat-inactivated for 20 min at 60 °C. In the case of BglII enzyme, heat-inactivation was not recommended, therefore, the reaction was stopped by separating endonucleases and DNA by gel electrophoresis. The DNA was, afterward, retrieved by gel extraction.

To remove small DNA fragments which remained after restriction digestion, digestion reactions were purified using QIAquick PCR Purification Kit.

Vector's and insert's adjacent DNA termini were joint together by DNA ligase. The following ligation reaction was prepared on ice using restriction digested DNA (Table 2.10):

Component	Reagent amount
T4 DNA ligase buffer (10x)	1 µl
Vector DNA	100 ng
Insert DNA	variable
T4 DNA ligase	1 µl
H_2O	up to 20 µl

Table 2.10: Components of ligation reaction

The amount of insert DNA used depended on the vector and insert of interest and was calculated using the following formula from the T4 DNA ligase protocol (Promega):

$$\text{insert DNA[ng]} = \frac{\text{vector[ng]} * \text{insert[kb]}}{\text{vector[kb]}} * \text{molar ratio of insert to vector} \quad (2.2)$$

The reaction was performed at room temperature for 10 min. T4 DNA ligase was then heat-inactivated for 10 min at 65 °C.

Preparation of *E. coli* for transformation

To prepare a stock of Top10 *E. coli*, bacterial cultures were inoculated from glycerol stocks into 3 ml LB medium. The culture was grown overnight at 37 °C with rigorous shaking (250 rpm). Following day, 1 µl of the culture was inoculated into 100 ml LB medium and grown at 37 °C with shaking until it reached an OD_{600} of 0.3-0.6. The bacteria were pelleted by centrifugation (10 min, 3000 rpm, 4 °C) and resuspended in 5 ml of LB mix for competent cells (1x LB, 10% PEG 8000, 5% DMSO, 10mM $MgCl_2$, 10mM $MgSO_4$, 10% Glycerol in H_2O). Aliquots of 200 µl were distributed into tubes pre-cooled on dry ice for cryopreservation of bacteria. The bacterial stocks were stored at -80 °C.

Transformation of Top10 *E. coli*

Chemically competence *E. coli* (strain Top10) were generated using $CaCl_2$. Top10 *E. coli*, were thawed on ice. 10 µl of ligation reaction was added to 100 µl KCM. 100 µl of Top10 *E. coli* were added to KCM mix (100mM KCl, 30mM $CaCl_2$, 50mM $MgCl_2$ in H_2O). The mix was incubated for 15 min on ice. DNA uptake was induced by heat shock for 90 s at 42 °C. To allow the bacteria to recover, 300 µl LB medium were added and the suspension was incubated for 1 h at 37 °C with shaking. After incubation, 250 µl of culture were streaked onto a selective LB agar plate and grown overnight at 37 °C .

Plasmid DNA purification

The EndoFree Plasmid Maxi Kit was used for purification of endotoxin-free plasmid DNA according to the manufacturer's instructions.

2.2.3 DNA and RNA isolation

TRIzol reagent for total RNA isolation

TRIzol reagent is single-phase solution of phenol and guanidine isothiocyanate which allows isolation of high-quality total RNA.

It was used according to the manufacturer's instructions. Briefly, cells were grown to 90 % confluency in 6 well plates, washed with PBS once and lysed with 1 ml TRIzol reagent. After incubation of 5 min at room temperature, 200 µl of chloroform were added. Centrifugation at 12 000 g for 15 min at 4 °C was performed to separate the phases into a lower organic phase, interphase and an upper aqueous phase containing the RNA. The aqueous phase was carefully transferred to a new reaction tube. To precipitate RNA, 10 µg RNase-free glycogen were added as carrier to the aqueous phase together with 500 µl of 100% isopropanol. Precipitated RNA was collected by centrifugation at 12 000 g for 10 min at 4 °C and the supernatant was removed without disturbing the RNA pellet. The pellet was washed with 75% ethanol. The RNA was resuspended in 30 µl of RNase-free H_2O preheated to 60 °C and stored at -80 °C.

mirVana Isolation Kit for total RNA isolation

The mirVana Kit combines organic extraction with acid-phenol:chloroform and solid-phase extraction thus ensuring superior yield and high purity. The procedure was performed as stated in the manufacturer's instructions for isolation of total RNA. The RNA isolated using mirVana kit was used in small RNA-sequencing (sRNA-seq) experiment.

DNase treatment of RNA samples

To remove DNA contamination in RNA samples, samples were treated with the RNAse-free DNase Treatment and Removal Reagents Kit. The procedure was performed in agreement with the manufacturer's instructions.

DNA isolation

The DNA, used in the methylation study, was isolated using DNeasy Blood & Tissue kit according to the manufacturer's instructions.

Determination of nucleic acid purity

The spectrophotometer NanoDrop 1000 was used to assess the purity of DNA and RNA samples. To assess the purity $A_{260/A280}$ and $A_{260/A230}$ ratios were used. $A_{260/A280}$ values of 1.8 for DNA and 2.0 for RNA were generally considered as pure and acceptable values of $A_{260/A230}$ were in range from 1.7 to 2.2[81].

RNA clean-up by 1-butanol treatment

Whenever the RNA sample showed presence of contaminants (e.g. chaotropic salt or organic solvent), it was further purified by treatment with 1-butanol and diethylether as described in Krebs et al[82]. The RNA sample volume was adjusted to 150 µl with RNase-free H_2O. 150 µl 1-butanol were added to the sample, the solution was vortexed and centrifuged for 3 min at 10 000 *g* to separate the phases. The upper organic layer was aspirated. 1-butanol extraction was repeated three more times. To remove residual 1-butanol from the sample, 500 µl diethylether was added to the RNA sample. Sample was centrifuged for 3 min at 10 000 *g*. The upper phase was aspirated. Diethylether extraction was repeated once.

Quantification of nucleic acids

Historically, concentration of DNA and RNA has been determined using spectrophotometry and measuring absorbance at 260 nm. As DNA, RNA and proteins all show maximal absorbance at 260 nm, they contribute equally to the measured absorbance of a sample. Therefore, spectrophotometers cannot distinguish between these macromolecules and quantification derived from absorbance measurement at 260 nm should be considered with caution. However, fluorescent-based quantification of nucleic acids provides the more sensitive and specific alternative[83]. This approach is based on measurement of fluorescence which is emitted upon binding of the dye specific for the nucleic acid of interest to the designated molecule. The Qubit 3.0 fluorometer was used together with the dsDNA BR Assay Kit or RNA BR Assay Kit to measure dsDNA and RNA concentrations, respectively. Quantification was performed in accordance with the manufacturer's instructions.

Determination of nucleic acid integrity

Agarose gel electrophoresis was performed to test the integrity of isolated DNA and RNA. For that purpose, 1% agarose gels stained with EtBr were used. Samples were separated by gel electrophoresis for 1 h at 100 V and the bands were visualized under transiluminator and documented using gel documentation E.A.S.Y. Doc Plus system.

2.2.4 Complementary DNA (cDNA) synthesis by reverse transcription reaction

Total RNA was reverse transcribed using miScript II RT Kit which allow the conversion of all RNA species into cDNA. Mature miRNAs and other non-coding RNAs (ncRNA) are polyadenylated by poly(A) polymerase and reverse transcribed into cDNA using oligo-dT primers. The used oligo-dT primers are composed of a 3' degenerate anchor and a universal tag sequence on the 5' end, thus allowing amplification of mature miRNA in the real-time PCR step. Usage of HiFlex buffer provides simultaneous conversion of mRNAs and other RNA species (including mature, pre-miRNAs, pri-miRNAs, ncRNAs and mRNA) into cDNA by using random nucleotide primers in addition to oligo-dT primers.

The reverse transcription reaction was prepared on ice as described below (Table 2.11).

Component	Reagent amount
HiFlex buffer (5x)	2 µl
Nucleics Mix (10x)	1 µl
Reverse Transcriptase Mix	1 µl
Template RNA	125 ng
H_2O	up to 10 µl

Table 2.11: Components of reverse transcription reaction contained in miScript II RT Kit

Reverse transcription was performed for 60 min at 37 °C, followed by heat-inactivation of reverse transcriptase for 5 min at 95 °C.

2.2.5 Quantification of relative miRNA/gene expression using qPCR

qPCR is a technique that allows precise detection and quantification of specific gene/miRNA target expression levels. The quantification can be achieved through usage of the fluorescent dye SYBR Green, which binds specifically to double-stranded DNA. Thus, an increase in fluorescence intensity is correlated with the copy number of the target cDNA.

The miScript SYBR Green PCR Kit (Qiagen) was used to quantify the expression of mature miRNAs and FastStart Universal SYBR Green Master Mix (Roche) was used for pre-miRNA, pri-miRNA or mRNA quantification. To normalize the variable amount of cDNA between samples, endogenous controls were run alongside specific gene/miRNA targets of interest. U6 small nuclear RNA (Rnu6) and small nucleolar RNA, C/D box 33 (Snord33) were used to normalize mature miRNA and the genes *Glyceraldehyde-3-phosphate dehydrogenase (Gapdh)*, *β-glucuronidase (Gusb)* and/or *TATA-box binding protein (Tbp)* served as endogenous controls for mRNA, pre-miRNA and pri-miRNA. Additionally, no-template controls were prepared for each primer pair as a control of cross-contamination. Each combination of cDNA and primer pair was analysed in triplicates in 10 µl reactions (Table 2.12).

Component	Reagent amount (miRNA)	Reagent amount (mRNA, pre- and pri-miRNA)
2x SYBR Green PCR Master Mix	5 µl	5 µl
10x miScript Universal Primer	1 µl	-
2 µM forward primer	1 µl	-
2 µM forward and reverse primer	-	1 µl
cDNA diluted 1:20	0.5 µl	-
cDNA undiluted	-	0.5 µl
H_2O	up to 10 µl	up to 10 µl

Table 2.12: Components of qPCR master mixes for miRNA; mRNA, pre- and pri-miRNA reactions

qPCR amplification reactions were executed in QuantStudio 7 Flex Real-Time PCR System (Thermo Fisher Scientific) using the run method listed in the Tables 2.13, 2.14.To validate the specificity of PCR, an amplicon melting curve was determined

for each reaction in every run. If melting curve was consisted of only one peak, the PCR reaction was judged as specific. Reactions with unspecific amplifications were excluded from further analysis.

	Step	Temperature	Time
	Initial Denaturation	95 °C	15 min
40 cycles	Denaturation	94 °C	15 s
	Primer annealing	55°C	30 s
	Elongation	70°C	30 s
	Final elongation	98 °C	10 min

Table 2.13: Run method of the QuantStudio 7 Flex Real-Time PCR-System used for miRNA quantification

	Step	Temperature	Time
	Initial Denaturation	95 °C	10 min
40 cycles	Denaturation	95 °C	15 s
	Primer annealing and elongation	60°C	1 min
	Final elongation	98 °C	10 min

Table 2.14: Run method of the QuantStudio 7 Flex Real-Time PCR-System used for mRNA, pre- and pri-miRNA quantification

Post-run data processing was conducted using LinRegPCR[84] (http://LinRegPCR.nl). Result of the LinRegPCR analysis are target gene's starting concentrations which can easily be compared between different samples. To calculate the concentrations LinRegPCR requires estimation of PCR efficiency, determination of the C_t value and corresponding fluorescence measurement at cycle C_t.

The basic equation for PCR kinetics states that the amount of amplicon after C_t cycles (N_c) is dependent on the starting concentration of amplicon (N_0) and amplification efficiency (E) (Equation 2.3). Amplification efficiency is defined as the fold increase of the DNA amount per cycle, ranging from 1 (no amplification) to 2 (complete doubling). Equation 2.3 can be rearranged to calculate the starting concentration 2.4.

$$N_c = N_0 * E^{c_t} \tag{2.3}$$

$$N_o = N_c * E^{c_t} \tag{2.4}$$

N_c value represent user-defined fluorescence threshold in the exponential phase of the reaction and C_t fractional number of cycles needed to reach that florescence threshold. Calculated N_0 value is expressed in arbitrary fluorescence units.

PCR efficiency calculation is based on a model, in which the cycle number is the independent variable and the log fluorescence value the dependent variable. The efficiency is derived from the slope of the regression line fitted to the data points of the model[85].

It was shown that the amplicon sequence is the main determinant of the PCR efficiency[86]. To reduce the variability of the PCR efficiency LinRegPCR averages the efficiency values of all reaction, in which the same aplicon was amplified[84].

Determination of the exponential phase of the PCR reaction is based on the regression model used to calculate the PCR efficiency. Data points on the regression line are referred to as data points in the Window-of-Linearity (W-o-L). For each amplicon group, a C_t value is set at 1 cycle below the top border of the W-o-L and the N_c value is determined as corresponding log fluorescence value for cycle C_t.

The relative expression of target gene/miRNA was calculated by averaging N_0 values of technical replicates. To be able to compare the N_0 value of target gene/miRNA across samples, the N_0 value of the target was normalized to the N_0 value of the endogenous control. The N_0 value of all samples are expressed relative to a randomly chosen control sample.

Data processing and statistical analysis were performed using R (https://www.R-project.org, v3.5.1). Normalised N_0 values were compared using an two-sided unpaired t-test. Differences were considered statistically significant if the p-value was ≤ 0.05.

2.2.6 Staining of paraffin-embedded HCC samples

Tumor and nodular samples (described in subsection 2.1.1) together with corresponding controls were embedded in paraffin by Dr. Abhishek Thavamani and Dr. Stefen Ohrnberger.

Immunostaining and Sirius Red staining of the paraffin-embedded samples was performed in the group of Prof. Jan Hengstler by Brigitte Begher-Tibble (Leibniz Research Centre for Environment and Human Factors, Dortmund).

Formalin-fixed and paraffin-embedded tissue samples were deparaffinized in Rotihistol and rehydrated in an ethanol gradient. The antigen retrieval step was performed using citrate buffer. For immunostaining, the MaxFlour 10/31 Immunofluorescence Detection Kit was used together with Early Growth Response (Egr1) and Alpha Smooth Muscle Actin (ACTA2) antibodies. The Egr1 and ACTA2 antibodies were diluted 1:800 and 1:100, respectively. Nuclei were stained using DAPI.

To visualise collagen depositions in the embedded tissue samples, Sirius Red staining was performed. Samples were de-waxed and hydrated. Nuclei were visualised by Weigert's haematoxylin. Subsequently, tissue slides were stained with Picro-Sirius Red and washed in acidified water. Samples were dehydrated in absolute ethanol, cleared in xylene and mounted in a resinous medium. Images were acquired by an Olympus confocal laser scanning microscope and quantified using ImageJ. The macro used for quantification is deposited in the GitHub repository (`https://ivanawinkler.github.io/mirna_paper/`). Measured stained percentage area of total area was compared between control, nodular and tumor samples using one-way ANOVA and a post-hoc Tukey test.

2.2.7 Chromatin immunoprecipitation (ChIP)

The used ChIP protocol is based on the procedure described by Daniel et al.[87], with some modifications.

To reversibly cross-link proteins bound to chromatin of the adherent cell lines, formaldehyde was directly added on the cell plate in a final concentration of 1 %. Per preparation (sample) 30 million cells were used. Added formaldehyde was diluted in PBS. Cells were incubated for 10 min and the cross-linking reaction was stopped by adding glycine to a final concentration of 0.125 M.

Cells were scraped off the plate in 1 ml of cell lysis/wash buffer (0.15 M NaCl, 0.005 M EDTA pH 7.5, 0.05 M Tris–HCl pH 7.5, 0.5 % NP40, dH_2O supplemented with protease inhibitor prior to use) and pelleted at 12,000 x g for 1 min. The pellet was resuspended in additional 1 ml of the cell lysis/wash buffer and passed through an insulin syringe. Following the centrifugation step, nuclei pellets were resuspended in 700 µl of nuclear lysis buffer (50 mM Tris-HCl pH 7.5, 1 % SDS, 20 mM EDTA).

Chromatin, which was released in the previous step, was subsequently fragmented by sonication into fragments of 200-1000 base pairs (bp). Sonication was performed for 15 pulses of 20 s with a 30 s resting interval at 4°C on high energy settings (Bioruptor, Diagenode). Efficiency of sonication was validated by loading a small aliquot of chromatin onto the agarose gel. Remaining chromatin was cleared by centrifugation at 12,000 x g at 4°C for 10 min.

10% of chromatin was used as input. Chromatin used in immunoprecipitation was diluted 10 times using ChIP Dilution Buffer (0.001 M EDTA, pH 8.0, 0.017 M Tris–HCl, pH 8.0, 0.01 % SDS, 1.1 % Triton-X 100, 0.17 M NaCl, dH_2O supplemented with protease inhibitor tablets prior to use).

Chromatin was split into two fractions. The first fraction was incubated with 2 μg antibody of interest (anti-PPARγ or anti-EGR1) and the second with IgG antibody used to assess the unspecific antibody binding. Samples were incubated overnight at 4 °C.

The following day, Protein A-coupled Dynabeads were blocked using Blocking buffer (0.1% polyvinylpyrrolidone (PVP), 0.1% ultraPure albumin (BSA) in PBST (0.1% TWEEN-20 in PBS)) for 30 min at room temperature.

The blocked bead suspension was added to the chromatin sample and chromatin-antibody-bead complexes were incubated for 6 h at 4 °C.

The samples were washed using cell lysis/wash buffer five times. Chromatin was eluted in 450 μl Elution buffer at 65°C (1 % SDS). To reverse the cross-linking, 20 μl 5M NaCl and 20 μl 0.5 M EDTA were added to all samples and samples were incubated at 65 °C overnight.

DNA was isolated using phenol:chloroform:isoamyl alcohol procedure. Briefly, phenol:chloroform:isoamyl alcohol (25:24:1) was added to precipitated and input fraction in final ration of aqueous and organic phase 1:1. Samples were briefly vortexed and centrifuged at 13000 rpm for 5 min. The upper phase was carefully transferred to a new tube and chloroform was added to the fractions in ratio of aqueous and organic phase of 1:1. Samples were vortexed and centrifuged at 13000 rpm for 5 min. The upper phase was carefully transferred to new tube. To precipitate the DNA, 3 V of ethanol, 0.1 V of 3 M sodium acetate, ph 5.2 and 20 μg of glycogen solution were added to all samples. Samples were incubated at -20°C for several hours. To pellet the DNA, samples were spun at 14 000 x g for 15 min at 4°C and washed with 70% ethanol. Finally, samples were resuspended in water.

Target protein-bound DNA in immunoprecipitated samples, together with input samples, was quantified in qPCR. Input fraction was quantified, alongside the im-

munprecipitated fraction to normalize for sources of variability, including amount of chromatin and DNA recovery. The Percent Input Method was used as method of normalisation. To calculate percent input following formula was used 2.5:

$$\%\ \text{of input} = 100 * 2^{((Ct_{input} - 6.644) - Ct_{IP})} \tag{2.5}$$

where Ct represents the intersection between an amplification curve and a threshold line as determined by QuantStudio Software software (version 1.3).

2.2.8 Small RNA-seq

To profile the whole miRNome, small RNA-sequencing (sRNA-seq) was performed on tumor and nodular tissue alongside the corresponding controls.

List of used samples and corresponding metadata are presented in Table 2.15.

RNA was isolated and DNase treated in accordance with the procedure described in section 2.2.3: mirVana Isolation Kit for total RNA isolation and Dnase treatment of RNA samples. RNA purity was determined using Nanodrop and RNA concentration was measured using Qubit 3.0 (section 2.2.3: Quantification of nucleic acids and Determination of nucleic acid purity). Integrity of the total RNA was controlled on an Agilent Technologies 2100 Bioanalyzer. Samples with RIN value $\geqslant 7$ were used in library preparation.

The RNA sequencing library were generated by Dr. Sabin Bhuju (Genome Analytics group, Helmholtz Centre for Infection Research, Braunschweig).

The library was generated from 1 µg of total RNA using TruSeq Small RNA Library Prep Kits v2 (Illumina) according to manufacturer's protocol. The libraries were sequenced on Illumina HiSeq2500 using TruSeq SBS Kit v3-HS (50 cycles, single ended run) with an average of 10^7 reads per RNA sample. FASTQ files were trimmed with Cutadapt (v1.11) removing Illumina RNA adapter sequences (TGGAATTCTCGGGTGCCAAGG) and nucleotides with PHRED scores below 20. For each FASTQ file a quality report was generated using FASTQC (`https://www.bioinformatics.babraham.ac.uk/projects/fastqc/`) (v0.11.4) tool before and after trimming. Alignment of trimmed FASTQ sequences was done using STAR Aligner (v2.5.2b)[88] against the mouse genome GRCm38 using miRBase (v21) annotation. One mismatch was allowed for successful alignment in at least 15 matches per sequence. Counts per smallRNA (feature)

Litter	Sample	Age/weeks	Gender	Liver-to-body-ratio	Sample type	Experiment
4	4-C	32	F	4.8	control	sRNA-seq, RNA-seq, proteomics
4	4-N	32	F	26.3	nodule	sRNA-seq, RNA-seq, proteomics
4	4-T	32	F	26.3	HCC_A	sRNA-seq, RNA-seq, proteomics
25	25-C	32	F	5.1	control	sRNA-seq, RNA-seq, proteomics
25	25-N	32	F	30.1	nodule	sRNA-seq, RNA-seq, proteomics
25	25-T	32	F	30.1	HCC_B	sRNA-seq, RNA-seq, proteomics
26	26-C	31	F	4.2	control	sRNA-seq, RNA-seq, proteomics
26	26-N	31	F	26.8	nodule	sRNA-seq
26	26-T	31	F	26.8	HCC_B	sRNA-seq, RNA-seq, proteomics
49	49-C	52	M	4.9	control	RNA-seq, proteomics
49	49-T	52	M	16.4	HCC_A	RNA-seq, proteomics
2	2-C	32	M	5	control	sRNA-seq
2	2-N	32	M	24.3	nodule	sRNA-seq, RNA-seq, proteomics
2	2-T	32	M	24.3	HCC_A	sRNA-seq, RNA-seq, proteomics
59	59-N	30	M	23.4	nodule	RNA-seq, proteomics
59	59-T1	30	M	23.4	HCC_A	RNA-seq, proteomics
59	59-T2	30	M	23.4	HCC_B	RNA-seq, proteomics
59	59-T3	30	M	23.4	HCC_B	RNA-seq, proteomics

Table 2.15: List of samples used in omics analysis and their corresponding meta-data

were calculated using STAR's parameter *quantMode* set to *GeneCounts*. Differential expression of small RNAs was determined using DESeq2[89] (v1.22.2).

The code for the bioinformatic analysis outlined here is available in the following url: `https://ivanawinkler.github.io/mirna_paper/`.

2.2.9 RNA-seq and proteomics

RNA was isolated using RNeasy Mini kit according to the manufacturer's instructions. List of the samples used for RNA-seq analysis and corresponding metadata are listed in Table 2.15.

The RNAseq library was generated using TruSeq RNA sample preparation kit according to manufacturer's protocol.

Preparation of RNA extracts and libraries was carried out by Dr. Abhishek Thavamani.

The libraries were sequenced at Genome Center of Max Planck Institute for Developmental Biology on Illumina HiSeq2000 device, producing 100 bp long, single ended read. On average 10^{*7} reads was sequenced per RNA sample.

FASTQ files were trimmed with Cutadapt[90](v1.11) removing Illumina RNA adapter sequences (AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC) and nucleotides with PHRED scores below 20. For each FASTQ file a quality report was generated by FASTQC (`https://www.bioinformatics.babraham.ac.uk/projects/fastqc/`) (v0.11.4) tool before and after trimming. Alignment of trimmed FASTQ sequences was done using STAR Aligner[88] (v2.5.2b) against the mouse genome GRCm38. Counts per gene (feature) were calculated using STAR's[88] parameter *quantMode* set to *GeneCounts*. Differential gene expression was determined using DESeq2[89] (v1.22.2).

The code for the bioinformatic analysis outlined here is available in the following url: `https://ivanawinkler.github.io/mirna_paper/`.

To prepare the protein extracts, flow-through of RNA lysate, obtained in first step of RNeasy preparation, was precipitated using acetone (2.5 V). Mixture was incubated overnight. Following day, proteins were precipitated by centrifugation at 14 000 rpm for 15 min. The pellet was resuspended in Urea buffer.

Protein extract was treated with 1 mM Dithiothreitol for 1h followed by 5.5mM Iodoacetamide treatment for an additional hour. Samples were then predigested with endoproteinase Lys-C for 3 h and digested with trypsin overnight. After digestion, peptides were purified on StageTips and measured on EASY-nLC II nano-LCs coupled

to an LTQ Orbitrap XL instrument. LC-MS/MS measurement and analysis of obtained data was performed by Dr. Katarina Matic. Preparation of protein extracts and protein sample processing was performed by Dr. Abhishek Thavamani.

2.2.10 Methylation analysis of miRNA promoters

Quantitative DNA methylation analysis was performed by matrix-assisted time-of-flight mass spectrometry (MassARRAY; Agena Bioscience) essentially as described previously[91] using primers listed in supplementary material. In brief, gDNA isolated from quiescent and activated pHSC was bisulfite-converted and used as template to generate PCR amplicons with a T7-promoter tag. *In vitro*-generated RNA from the amplicons was cleaved by RNase A providing specific fragments which were analysed by MassARRAY. Fragments with ambiguous or too low ($\leq$1500) or too high ($\geq$7000) masses were omitted from the analysis.

MassARRAY analysis was performed by Dr. Dieter Weichenhan (Epigenomics and Cancer Risk Factors division, DKFZ Heidelberg).

2.2.11 Luciferase assay

To experimentally validate the functionality of predicted miRNA targeting, a luciferase gene reporter assay was used. The predicted target site or the full length 3'-UTR and a mutated version were subcloned downstream of the luciferase gene and transfected together with a specific miRNA mimic.

The list of miRNA:mRNA pairs for which targeting was experimentally validated and corresponding experimental data are presented in Table 2.16.

pmirGLO Dual-Luciferase miRNA Target Expression Vector was used as backbone vector to clone in a partial or full length 3'-UTR region of *Col1a1, Pdgfa, Adamts15* and *Tgfbr1* genes. Following vector transfection, cells were transfected with a miRNA mimic of interest. Final concentration of all miRNA mimics was 50 μM. Cloning and transfection procedure were performed as described in sections 2.2.2 and 2.2.1: Cell transfection, respectively.

To mutate miRNA sites in all pmiRGLO vectors, Q5 Site-Directed Mutagenesis kit was used according to the manufacturer's instructions.

Cloning and mutagenesis procedure was performed by Catrin Bitter (AG Nordheim).

After 24 h, luciferase activity was measured using Dual-Luciferase Reporter Assay (Promega) according to the manufacturer's instructions. Luciferase activity was measured using OPTIMA FluoroSTAR. To compare measured luciferase fluorescence units (LFU), one-way ANOVA and a post-hoc Tukey test were used.

Primers used for cloning and mutagenesis are listed in Table 2.2 and used miRNA mimics and their sequence are shown in Table 2.17.

miRNA mimic	Sequence	Identifer
mmu-miR-29c-3p	UAGCACCAUUUGAAAUCGGUUA	Dharmacon, C-310522-07
mmu-miR-let-7a-5p	UGAGGUAGUAGGUUGUAUAGUU	Dharmacon, C-310502-07
mmu-miR-let-7c-5p	UGAGGUAGUAGGUUGUAUGGUU	Dharmacon, C-310505-07
mmu-miR-let-7g-5p	UGAGGUAGUAGUUUGUACAGUU	Dharmacon, C-310374-07
mmu-miR-338-3p	UCCAGCAUCAGUGAUUUUGUUG	Dharmacon, C-310547-06

Table 2.17: List of used miRNA mimics, their sequence and source

miRNA	Target mRNA	Experiment
mmu-miR-29c-3p	*Col1a1*	Luciferase assay, qPCR
mmu-miR-29c-3p	*Pdgfa*	Luciferase assay, qPCR
mmu-miR-29c-3p	*Pdgfb*	qPCR
mmu-miR-29c-3p	*Tpm1*	qPCR
mmu-miR-29c-3p	*Col1a2*	qPCR
mmu-miR-29c-3p	*Col4a2*	qPCR
mmu-miR-29c-3p	*Col4a5*	qPCR
mmu-miR-29c-3p	*Col5a2*	qPCR
mmu-miR-29c-3p	*Adamts14*	qPCR
mmu-miR-29c-3p	*Lamc1*	qPCR
mmu-let-7a-5p, mmu-let-7c-5p, mmu-let-7g-5p	*Col1a1*	Luciferase assay, qPCR
mmu-let-7a-5p, mmu-let-7c-5p, mmu-let-7g-5p	*Adamts15*	Luciferase assay, qPCR
mmu-let-7a-5p, mmu-let-7c-5p, mmu-let-7g-5p	*Tgfbr1*	Luciferase assay, qPCR
mmu-let-7a-5p, mmu-let-7c-5p, mmu-let-7g-5p	*Col4a2*	qPCR
mmu-let-7a-5p, mmu-let-7c-5p, mmu-let-7g-5p	*Col4a5*	qPCR
mmu-let-7a-5p, mmu-let-7c-5p, mmu-let-7g-5p	*Col5a2*	qPCR
mmu-let-7a-5p, mmu-let-7c-5p, mmu-let-7g-5p	*Col1a2*	qPCR
mmu-let-7a-5p, mmu-let-7c-5p, mmu-let-7g-5p	*Adamts14*	qPCR
mmu-let-7a-5p, mmu-let-7c-5p, mmu-let-7g-5p	*Loxl4*	qPCR
mmu-miR-338-3p	*Col1a1*	Luciferase assay
mmu-miR-338-3p	*Adamts15*	Luciferase assay

Table 2.16: List of miRNA:mRNA pairs for which targeting was experimentally validated

2.2.12 Modulation of target gene expression with miRNA mimic

Monitoring of the target gene expression upon miRNA mimics transfection was used as alternative approach to experimentally validate the functionality of predicted miRNA targeting (Table 2.16. miRNA mimics are double-stranded RNA oligonucleotides designed to mimic the function of endogenous, mature microRNAs. miRNA mimics were transfected as described in section 2.2.1: Cell transfection. Gene expression was quantified as reported in section 2.2.5.

2.2.13 miRNA target prediction

To identify miRNA candidates with potential contributions to mHCC formation, four filtering criteria were employed: (i) differential expression in tumours versus controls ($p_{adj} \leq 0.05$), (ii) sufficient expression level ($\geq$ 10 counts across all samples), (iii) miRNA conservation (mice and humans) and (iv) similar expression pattern of at least one family member in humanHCCs (hHCCs) ($\geq$ 10% of cases in The Cancer Genome Atlas's (TCGA's) cohort) (Figure 2.1).

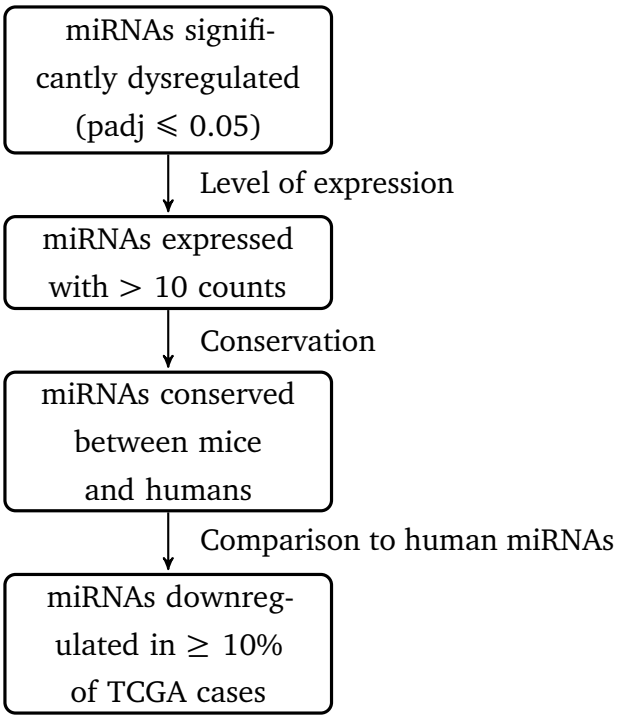

Figure 2.1: The procedure of miRNA candidate selection. The aim of the selection procedure is to profile miRNA candidates with contribution to carcinogenesis. To shortlist the final candidates, miRNA were filtered for level of expression, conservation and comparison to human miRNAs.

As it is more likely that miRNAs with higher absolute expression levels will have a more significant biological impact upon dysregulation, to profile the miRNA candidates, only those miRNAs expressed at the level of 10 counts across all samples were used in further analysis.

Second filtering criteria relied on miRNA conservation. As the final goal of the study is to translate conclusions gained from animal model to human patients, miRNAs that were used in target identification were filtered for conservation between mice and humans.

Additional level of filtering involved retention of those miRNAs, which showed similar expression pattern of at least one family member in at least 10% of human HCC from The Cancer Genome Atlas's (TCGA's) HCC cohort.

Subsequent to identification of tumour-associated miRNAs, screening of potential miRNA targets was performed using the DIANA microT-CDS (v5) [92,93] and TargetScan (v7.2) [14,94] databases. To increase the accuracy of downregulated miRNA target predictions, RNA-sequencing (RNA-seq) and proteomics analyses were performed on largely overlapping samples as used for sRNA-seq (Subsection 2.2.9) and target mRNA candidates generated through bioinformatic analysis were correlated with the upregulated genes found by RNA-seq and proteomics. Similarly, upregulated miRNA targets were identified.

To profile evolutionary conserved miRNA targeting, an analysis comparable to the above was performed on the TCGA datasets of hHCCs. Subsequently, miRNA:target mRNA pairs of both murine and human datasets were overlapped and only conserved pairs were further used in gene set enrichment analysis (GSEA), utilising KEGG (v6.2) [95] and Reactome [96] (v67) pathways (Figures 2.2 and 2.3). Filtering for evolutionary conserved targeting was performed only for miRNA:target mRNA subsetted using mouse RNA-seq data. miRNA:target mRNA pairs which were subsetted using proteomics dataset were not compared to hHCC and not filtered for evolutionary conserved miRNA targeting.

To perform gene enrichment analysis, overrepresentation analysis using a hypergeometric test was performed.

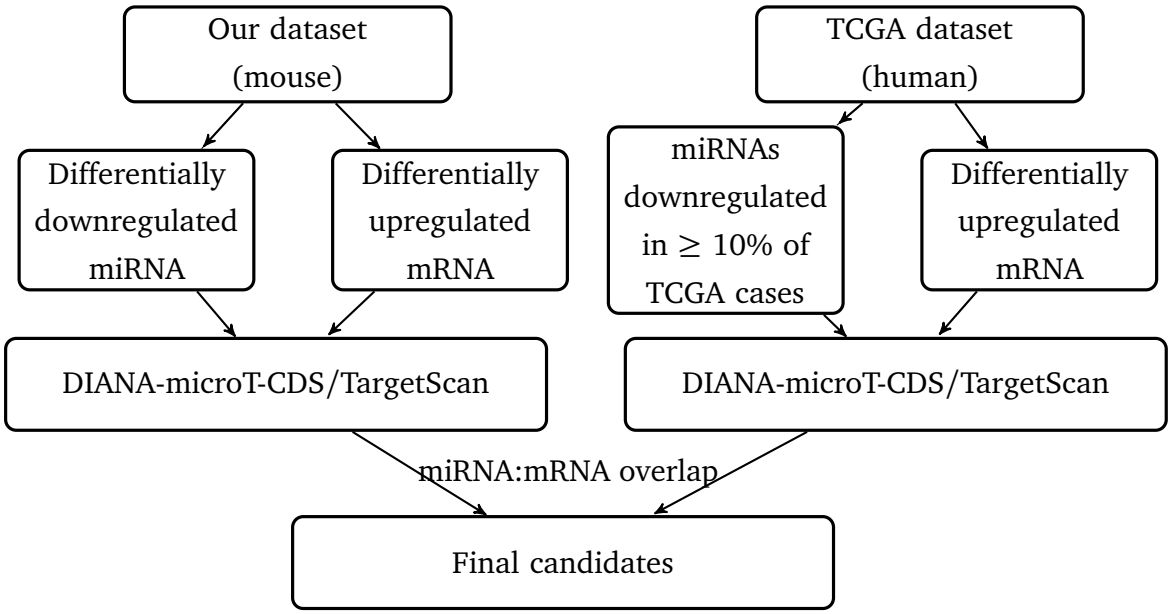

Figure 2.2: Target identification procedure for downregulated miRNAs. To identify targets of downregulated miRNAs, target mRNA candidates generated through bioinformatic analysis were correlated with the upregulated genes found by RNA-seq/proteomics. To profile evolutionary conserved targeting same analysis was done for human HCC TCGA dataset. miRNA:target mRNA pairs of both dataset were overlapped and only those pairs that showed conservation were used in gene enrichment analysis. Mouse miRNA:target mRNA pairs profiled using proteomics dataset were not overlapped with human dataset.

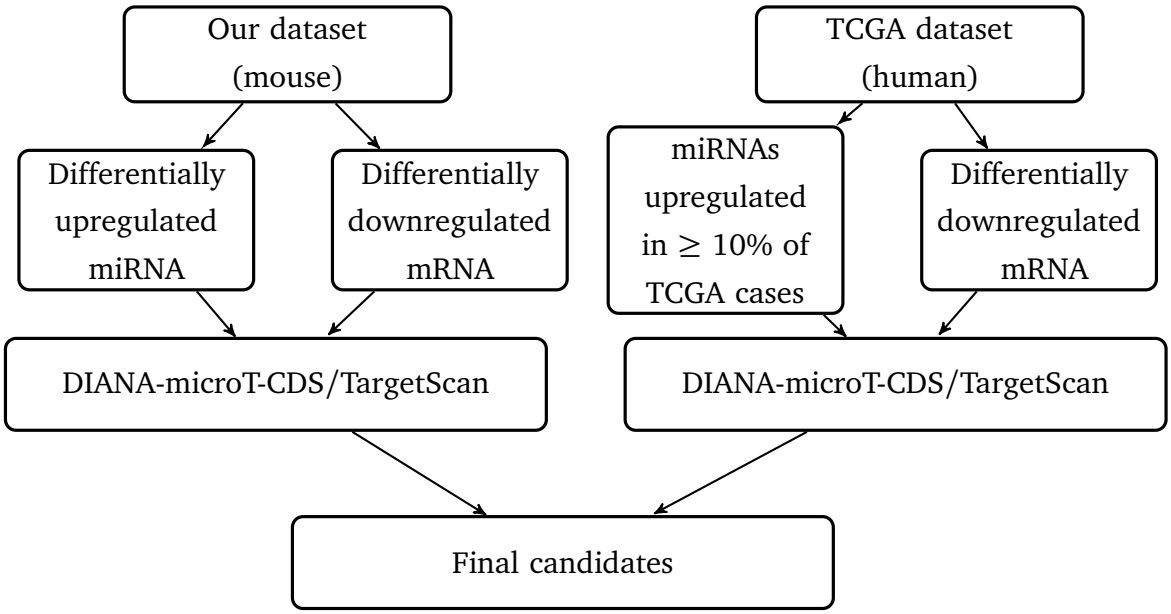

Figure 2.3: Target identification procedure for upregulated miRNAs. To identify downregulated genes targeted by upregulated miRNAs similar approach was used but without filtering step for conserved miRNA:mRNA pairs. As upregulated miRNAs in both human and mouse dataset globally downregulate the same targets but miRNA:target mRNA pairs themselves are not conserved, target genes dowregulated in at least 15% of human tumor samples were considered as candidate genes. Mouse miRNA:target mRNA pairs profiled using proteomics dataset were not overlapped with human dataset.

2.2.14 miRNA transcription start site prediction

A bioinformatic approach was used to identify transcription start sites (TSSs) of miRNAs of interest. As pri-miRNA are rapidly cleaved to pre-miRNA mapping miRNA transcription start sites using conventional TSSs mapping approaches is highly challenging[97]. Global nuclear run-on sequencing (GRO-seq) is a technique used to quantifying nascent transcripts. Briefly, transcription is halted, nuclei are isolated and transcription briefly restarted using labelled nucleotides resulting in labelled RNA molecules. Labelled RNA molecules are precipitated and sequenced[98]. GRO-seq shows sharp peaks around TSSs in both the sense and antisense directions and continuous signal of lower intensity throughout the entire transcript allowing to map TSSs of very transient transcripts. To map miRNA TSSs, six human (SRR014283, SRR574824, SRR1015583, SRR1145822, SRR1745515, SRR2961002) and nine mouse (SSR097858, SRR097863, SRR097864, SRR1517780, SRR1772450, SRR1991266, SRR3051599, SRR3051601,

SRR5816144) GRO-seq datasets deposited in the Gene Expression Omnibus (GEO) were used. To download the datasets the FASTQdump (v2.8.2) tool was used, FASTQC (v0.11.4) tool to examine the quality of the datasets, STAR (v2.5.2b) to perform the read alignment[99] and HOMER (v4.10) to perform peak analysis[100]. Integration of miRNA TSSs from all datasets resulted in the list of unique miRNA TSSs.

2.2.15 Transcription factor binding site prediction

As promoter regions of miRNAs, regions 1000 bp downstream of a TSS and 500 bp upstream of a TSS were considered. To identify transcription factors which can potentially bind to the miRNA promoters, FIMO (v4.12.0), part of the MEME suite, was used[101]. FIMO is position weighted matrix based tool and it scans provided sequences for individual matches to each of the used motifs. JASPAR CORE (v7), which contains a curated, non-redundant set of profiles derived from published and experimentally defined transcription factor binding sites for eukaryotes, was used as motif database[102]. To refine the prediction two RNA-seq datasets (GSE78853 - inactive (vehicle treated) and activated (TGF-β treated) hepatic stellate cells and GSE93313 - quiescent and inactive pancreatic stellate cell lines) were used. Dataset GSE78853 was downloaded as raw reads and processed using DESeq2 (v1.22.2), while dataset GSE93313 was downloaded as table of differentially expressed genes. The code for the bioinformatic analysis outlined here is available in the following url: `https://ivanawinkler.github.io/mirna_paper/`.

2.2.16 TCGA data processing

The data used for this analysis were generated by the TCGA Research Network: `http://cancergenome.nih.gov/`. The TCGA project characterized over 20,000 primary cancers generating over 2.5 petabytes of genomic, epigenomic and transcriptomic data. TCGA's sRNA-seq, RNA-seq, methylation data and copy number variation (CNV) data for liver HCC, lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC) and breast invasive carcinoma (BRCA) together with the corresponding metadata were downloaded from the TCGA data portal.

sRNA-seq and RNA-seq data were downloaded as raw reads and processed further using DESeq2 (v1.22.2). As not every tumor sample has a corresponding control sample, normalised reads of control samples were averaged and reads of individual

tumor samples were compared to this average control value. To generate normalised reads, DESeq2 package was used.

Copy number variation (CNV) probe values were mapped to genes and normalised to the value of 2, which was taken as reference gene copy number. To process CNV data, R package CNTools was used.

Beta values of probes in the methylation dataset were mapped to the genes. Beta values of control samples were averaged and compared to beta values of individual tumor samples.

The code for the bioinformatic analysis outlined here is available in the following url: `https://ivanawinkler.github.io/mirna_paper/`.

2.2.17 Analysis of microRNA:mRNA interactions across fibrotic cancers

To assess if a subset of miRNAs regulates common mRNAs across diverse cancer types, the multivariate linear regression approach introduced by *Jacobsen et al.* [103] was used.

For every pair (k,l) of mRNA k and miRNA l of interest and a given tumor type, the following regression model was employed. Let $y = (y_i) \in \mathbf{R}^n, i = 1,...,n$ be the expression of mRNA k for all n samples of a given cancer type. y is then modeled as the dependent variable in a linear regression with CNV intensities, $x_{cnv} = (x_{cnv,i})$, DNA methylation beta fold changes, $x_{me} = (x_{me,i})$ and expression of miRNA l, $x_l = (x_{l,i})$ as independent variables where $i = 1,...,n$ again denote the n tumor samples of a given cancer type:

$$y_i = \beta_0 + \beta_{cn} x_{cn,i} + \beta_{me} x_{me,i} + \beta_l x_{l,i} \quad , \quad i = 1,..,n \tag{2.6}$$

where β_0 is the intercept, β_{cn}, β_{me} and β_l are the regression coefficients for the CNV, methylation and miRNA expression covariates, respectively. The model was fitted using the standard assumptions of multivariate regression with the Python (v3.6.7) package *statsmodels* (v0.9.0).

Processing of the TCGA data used to fit the model is described in section 2.2.16.

The rank-statistic approach described in [103] was used to evaluate the relative strength of miRNA:mRNA association across the different cancer types, giving an indication of consistent dysregulation patterns of miRNA:mRNA pairs.

The code for the bioinformatic analysis outlined here is available in the following url: `https://ivanawinkler.github.io/mirna_paper/`.

Chapter 3

Results

3.1 sRNA dysregulation intensifies with tumor progression

To investigate the global contributions of miRNAs to the aforementioned processes, small RNA-sequencing (sRNA-seq) analysis of nodular and tumor samples derived from livers of *SRF-VP16*^{iHep} mice and corresponding controls was performed. sRNA-seq analysis was performed on four control, four nodular and four tumor samples.

FASTQ files obtained in the sequencing runs were quality-controlled using FASTQC tool. Obtained reads of all samples showed consistently high quality along the full read length (average PHRED score of all nucleotide positions across all sequenced reads and samples ≥ 30).

Sequencing reads were aligned against the mouse genome GRCm38 using STAR[88]. On average 64.67% of all reads across all samples was uniquely mapped and used in differential miRNA expression (DME) analysis.

DME was performed using DESeq2. Typically, controlled experiments involving RNA-seq or proteomics are performed on smaller number of replicas (three to four replicas per group). Such experimental set-up tends to result in highly variable dispersion estimates (i.e. intra-group variability) for each gene. DESeq2 assumes that genes of similar expression level share similar dispersion and it firstly estimates gene-wise dispersions using maximum likelihood. Next, the algorithm determines the local parameter of the estimates distributions and shrinks the gene-wise estimates towards the predicted values. Figure 3.1(A, C and D) shows normalised reads obtained in sRNA-seq before and after the shrinkage toward predicted dispersion values. This approach helps

in avoiding false positives, as it ensures that genes represented with the low counts do not appear significantly dysregulated[89]. RNA-seq and sRNA-seq data occasionally identify isolated instances of very large counts which appear unrelated to the study design and which are considered as outliers. To identify the genes that show consistent behaviour across all samples in particular group in contrast to outliers DESeq2 uses Cook's distances calculation. Cook's distance is a value which shows how much each samples is influencing the fitted coefficient for each gene[89]. Figure 3.1B shows average Cook's distance for each gene in all samples used in the sRNA-seq data. The median of Cook's distance is comparable in all samples used in sRNA-seq, indicating no outliers.

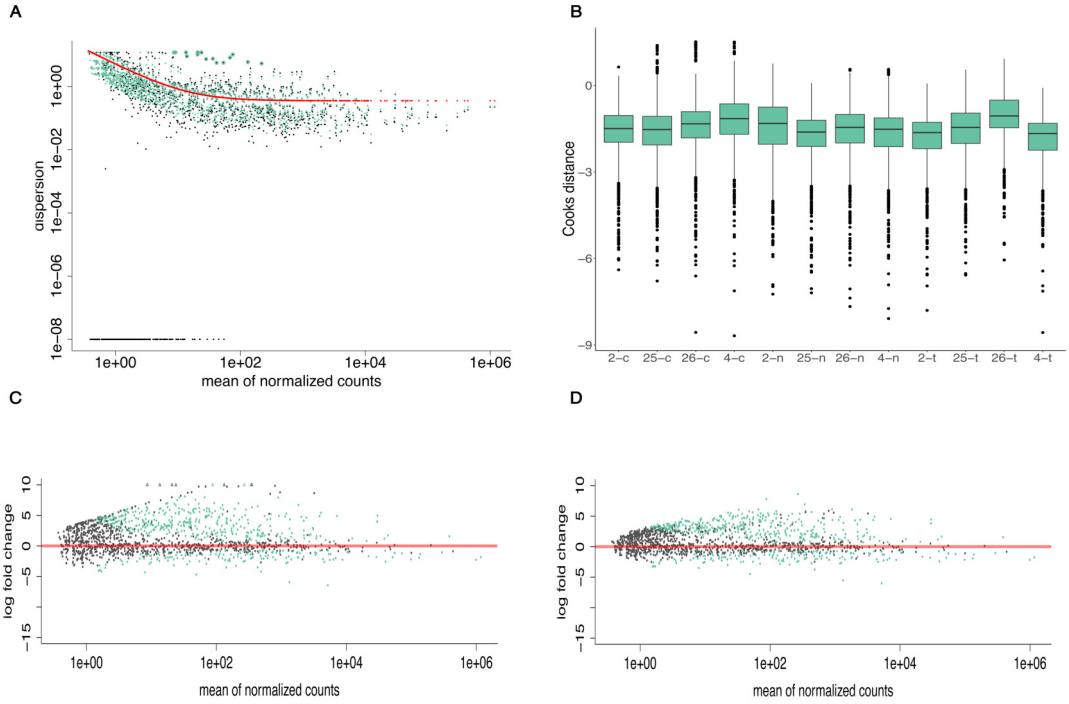

Figure 3.1: Dispersion plots and outlier plots indicate small number of outliers in sRNA-seq dataset. (A) Dispersion plot of mean normalized counts of all detected genes across all samples. Gene-wise estimates are shown in black, fitted values in red and the final maximum *a posteriori* estimates used in testing in green. (B) Cook distances of all samples used in sRNA-seq. (C-D) MA plots showing $\log_2$-fold changes of the treatment (tumor and nodules compared to control) over the mean of normalized reads. (C) shows unshrunken $\log_2$-fold changes and (D) shows shrinkage of genes $\log_2$-fold changes as a result of the incorporation of zero-centered normal prior.

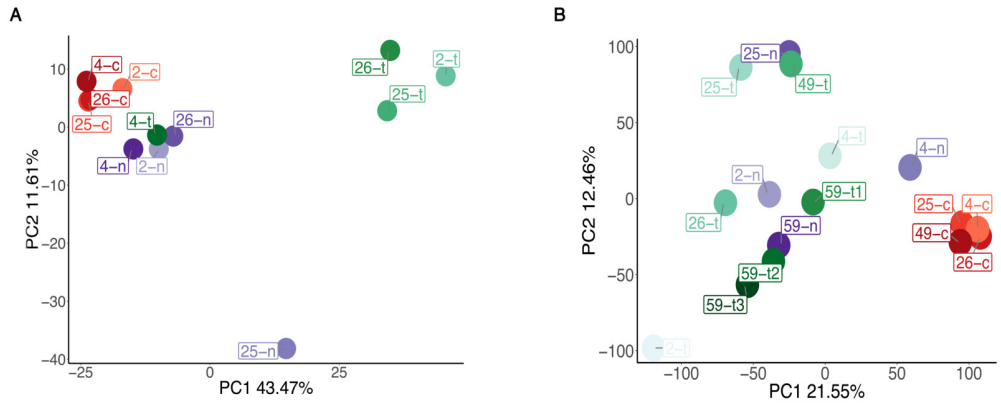

Figure 3.2: Principal component analysis (PCA) plot shows that sRNA dysregulation intensifies with tumor progression. (A) PCA plot of sRNA-seq dataset. (B) PCA plot of RNA-seq dataset. PC1 represent principal component 1 and PC2 principal component 2. Control samples are shown in red, nodular samples in green and tumor samples in purple colors.

To identify miRNA candidates with potential contributions to mHCC formation and microenvironment development, four filtering criteria were applied: (i) differential expression in tumors versus controls ($p_{adj} \leq 0.05$), (ii) sufficient expression level ($\geq$ 10 counts across all samples), (iii) miRNA conservation (mice versus humans) and (iv) similar expression pattern of at least one miRNA family member to hHCCs ($\geq$ 10% of cases in TCGA's cohort). Criteria iii and iv were applied to identify conserved miRNA expressions and functions with the final aim to compare the conclusions gained from our animal model to human patients. Upon applying these criteria, 52 miRNAs were found to be significantly downregulated and 31 significantly upregulated. These mouse miRNAs are dysregulated in at least 10% of human TCGA cases ($\geq$ 1.5-fold).

To increase the accuracy of miRNA target predictions, RNA-sequencing (RNA-seq) analyses was performed on largely overlapping samples as used for sRNA-seq (Table 2.15).

Principal component analysis (PCA) of obtained sRNA-seq data reveals that with the tumor progression, miRNA dysregulation intensifies. While nodular samples cluster closer to controls, tumors show expression profiles that are distinctly different from control samples, with the exception of tumor 4-t which clusters with the nodules. In contrast, PCA of RNA-seq data shows that nodular samples cluster with tumor samples. Although, not all nodular samples have corresponding tumor samples, in general, nod-

ules and tumors originating from the same livers cluster together, indicating partially tumor-specific differences in gene expression profiles.

3.2 Downregulated miRNAs regulate genes enriched in extracellular matrix pathways while upregulated miRNAs genes enriched in various metabolic pathways

Subsequent to identification of tumor-associated miRNAs, the screening of potential miRNA targets was performed using the DIANA microT-CDS[92,93] and TargetScan[14,94] databases. To identify targets of downregulated miRNAs, target mRNA candidates generated through bioinformatic analysis were overlapped with the upregulated genes found by RNA-seq and mass spectrometry-based proteomics analysis. Similarly, targets of upregulated miRNAs were identified.

To profile evolutionary conserved miRNA targeting, an analysis comparable to the above was performed on the TCGA datasets of hHCCs. Subsequently, miRNA:target mRNA pairs of both murine and human datasets were overlapped and only conserved pairs were further used in GSEA, utilising KEGG[95] and Reactome[96] pathways.

GSEA analysis of target genes of downregulated miRNAs profiled using RNA-seq showed strong over-representation of proteins involved in ECM function, integrin signaling and Rho GTPase-related pathways (Figure 3.3A) indicating the importance of miRNAs in regulation of the fibrotic microenvironment in HCC development. GSEA analysis of target genes of downregulated miRNAs profiled using proteomics dataset showed strong over-representation of Rho GTPase-related pathways while proteins involved in ECM function and integrin signaling were not detected in the proteomics analysis (Figure 3.3C).

Target prediction analysis based on integrated comparison to hHCC and filtering for evolutionary conserved targeting performed using RNA-seq dataset profiled 783 target genes of downregulated miRNAs and 3058 target genes of upregulated miRNAs. In contrast, target prediction analysis without integrated comparison to hHCC using proteomics dataset profiled 49 target genes of downregulated miRNAs and 16 of upregulated miRNAs. Overlap between target genes of downregulated and upregulated miRNAs subsetted using RNA-seq and proteomics is 11 and 14, respectively. In general,

number of genes/proteins identified in proteomics dataset in comparison to RNA-seq is noticeably smaller leading to significantly smaller number of identified miRNA target genes using proteomics dataset. Hence, larger number of pathways identified in the GSEA using RNA-seq-subsetted target genes is a result of larger number of identified target genes using this approach.

GSEA analysis of target genes of upregulated miRNAs derived using RNA-seq and proteomics showed strong over-representation of metabolic genes. Pathways identified to be regulated by upregulated miRNA include amino acids-, lipid-, fatty acids-, pyruvate- and Citric Acid (TCA) cycle-related pathways (Figure 3.3B, D). As in the case of target genes of downregulated miRNA, GSEA analysis of target genes subsetted using RNA-seq approach yielded more pathways in comparison to proteomics approach.

The remainder of this study focuses experimentally primarily on the downregulated miRNA profiled to target ECM- and Rho GTPase-related pathways.

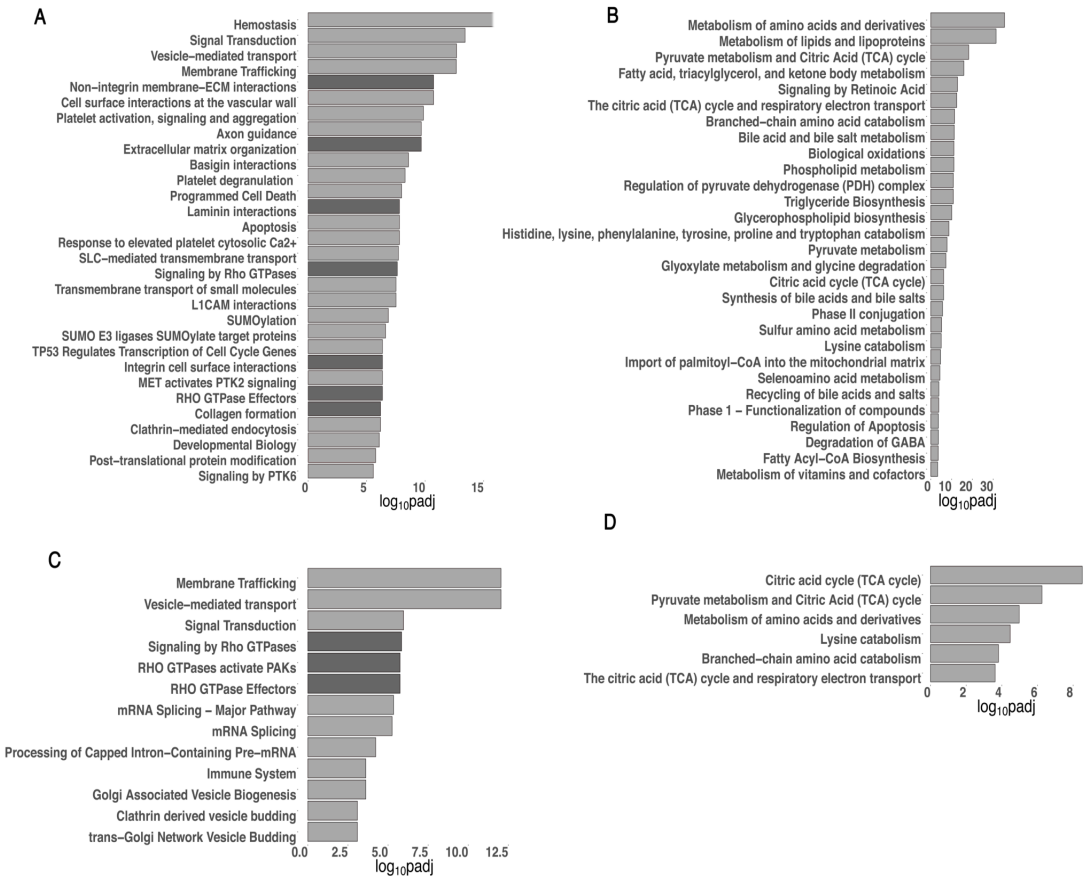

Figure 3.3: Downregulated miRNAs regulate genes enriched in extracellular matrix pathways while upregulated miRNAs genes enriched in various metabolic pathways. (A) Gene set enrichment analysis of target genes of downregulated miRNAs detected in RNA-seq and found to be conserved in murine (*SRF-VP16*^{iHep}) and human (TCGA) HCCs. Pathways depicted by bars in dark grey represent ECM- and Rho GTPase-related pathways. (B) Gene set enrichment analysis of target genes of upregulated miRNAs detected in RNA-seq. (C) Gene set enrichment analysis of target genes of downregulated miRNAs detected in mass spectrometry-based proteomics analysis. (D) Gene set enrichment analysis of target genes of upregulated miRNAs detected in mass spectrometry-based proteomics analysis.

3.3 A subset of miRNAs targets ECM-linked and fibrosis-associated genes in mHCC

Excessive accumulation of ECM leads to fibrosis development and promotes tumorigenesis by supporting enhanced growth, survival and proliferation of tumor cells[20].

To evaluate the progression of fibrosis alongside tumor progression, Sirius Red and alpha-smooth muscle actin (Acta2) staining of *SRF-VP16*^{iHep} pre-cancerous nodular and HCC tissue was performed, thereby quantifying fibrotic collagen depositions and HSC activation[14], respectively. This analysis reveals - in correlation with tumor progression - a gradual increase in both area covered by collagen (Figure 3.4A, C) and density of aHSCs (Figure 3.4B, D).

To demonstrate that HSC activation occurs in the vicinity of SRF-VP16-expressing hepatocytes, Early growth response 1 (Egr1) immunostaining was performed. As target gene of SRF[104], Egr1 is expressed at low levels in control samples, while its expression is high in SRF-VP16 nodular and tumor samples. Co-localisation of Acta2 and Egr1 signals confirms that aHSCs are part of the tumor microenvironment, providing further support to the significance of fibrosis in SRF-VP16-driven tumor development (Figure 3.4).

As the fibrotic microenvironment is more strongly developed in tumor tissue than nodular tissue (Figure 3.4B), the remainder of this study focuses on miRNA dysregulation in tumors.

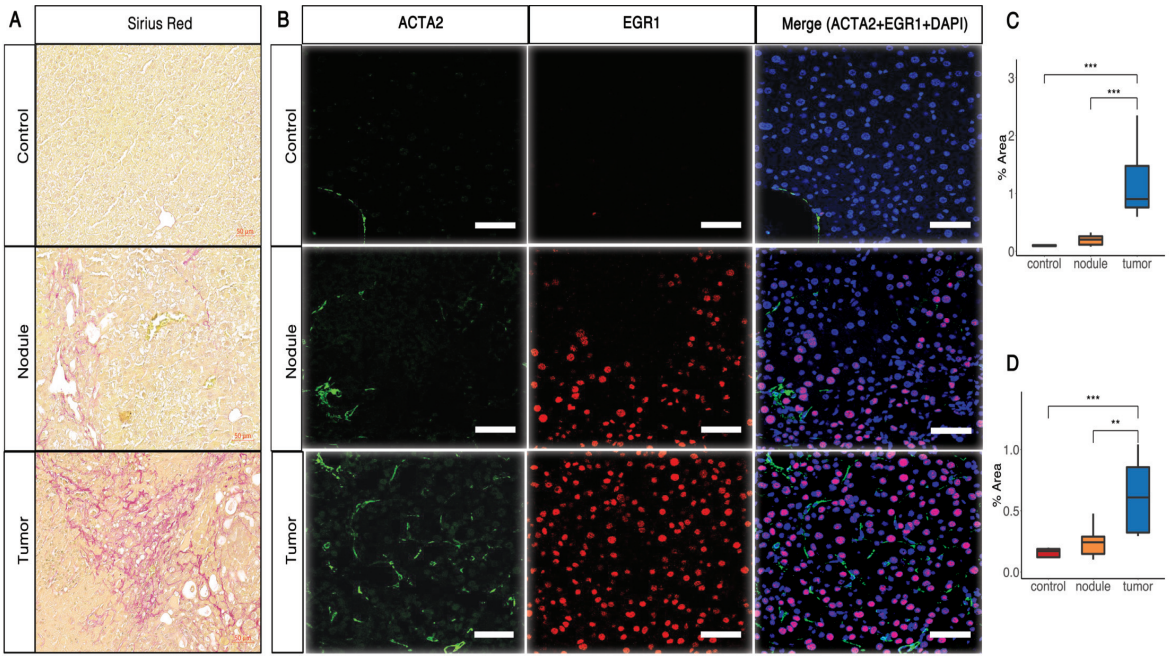

Figure 3.4: SRF-VP16-driven tumors progression is associated with fibrotic microenvironment development. (A) Sirius Red staining of control, nodular and tumor liver samples isolated from *SRF-VP16*^{iHep} mice. Scale 50 µm. (B) Acta2 (aHSCs), Egr1 (SRF-VP16 expressing hepatocytes) and merge (all cells) of control, nodular and tumor liver samples isolated from *SRF-VP16*^{iHep} mice. Scale 50 µm. (C) Quantification of Sirius Red signal shown in A. (D) Quantification of Acta2 signal shown in B. Data are shown as median, first and third quartile ("hinges") and 95% confidence interval of median ("notches"). * p-value $\leq$ 0.05, ** p-value $\leq$ 0.01, *** p-value $\leq$ 0.001. Images were acquired by Brigitte Begher-Tibbe (AG Hengstler).

Upon dissection of the ECM- and Rho GTPase-related pathways to individual genes and their associated regulatory miRNAs, constructs a miRNA:target mRNA network (Figure 3.5), comprised of a subset of miRNAs downregulated in tumors. To identify miRNA hubs in the network, all miRNAs which regulate ECM-related genes were filtered for the number of target genes and network coverage. Only miRNAs which target at least eight ECM-related genes and miRNAs which provide maximum targeting coverage of the genes in the network were retained. This miRNA network reveals that the miRNA families miR-30 and let-7, together with miR-335, miR-338 and miR-29c, control a majority of genes related to ECM function, as well as, integrin and Rho signaling. For simplicity, this subset of miRNAs is referred to as potential Anti-Fibrotic miRNAs (AF-miRNAs).

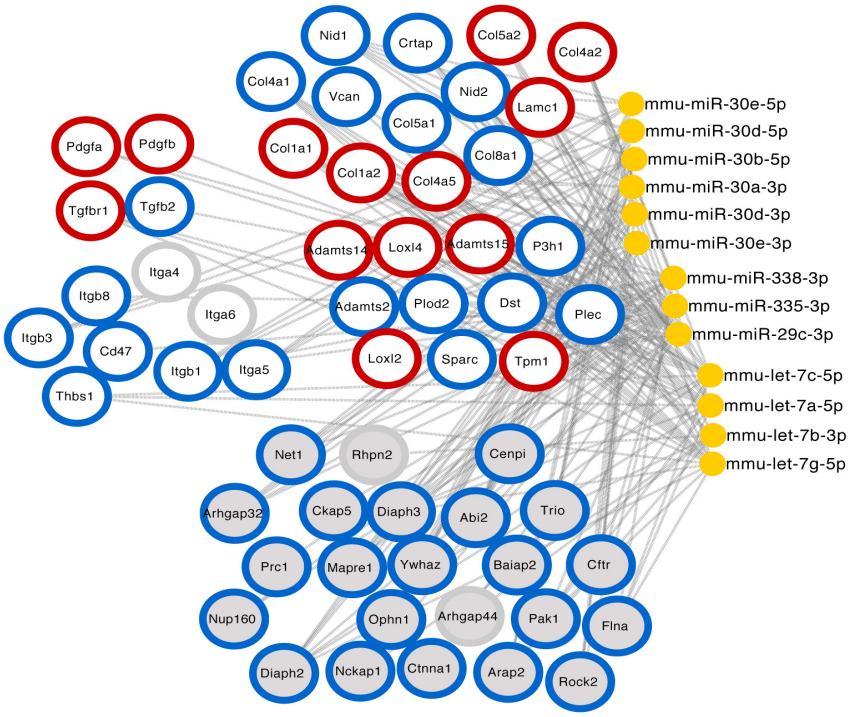

Figure 3.5: A subset of miRNAs targets ECM-linked and fibrosis-associated genes in mHCC. Network of miRNA:mRNA pairs, which encompasses miRNA regulation of genes contributing to the pathways highlighted in dark grey in Figure 3.3. Genes are grouped in structural (upper right), signaling (upper left) and remodeling (middle right) components of the ECM, as well as genes related to integrin signaling (middle left) and Rho signaling (bottom). Rims of gene nodes: red (ECM-related genes characterized in this study and targeted by here characterized miRNAs), blue (genes not characterized in this study but targeted by here characterized miRNAs), grey (genes not characterized in this study and targeted by here non-characterized miRNAs). Here characterized miRNAs (mmu-miR-30e-5p, mmu-miR-30d-5p, mmu-miR-338-3p, mmu-miR-335-3p, mmu-miR-29c-3p, mmu-let-7a-5p, mmu-let-7c-5p and mmu-let-7g-5p) target all ECM-related proteins of the network. Here characterized genes (red rims) represent important structural, remodeling and signaling component of the ECM.

Collagens represent the most dominant structural proteins of the ECM[105]. Most collagen family members in the network are regulated by the let-7 family and miR-29c, with a contribution of miR-335 and miR-338, while laminin gamma 1 (*Lamc1*) is targeted by miR-29c (Fig. 1B). ECM remodeling components, i.e. A Disintegrin and Metalloproteinases with a ThromboSpondin motif (*Adamts*) and LOX-like (*Loxl*) family members are regulated by the let-7 family, miR-29c and miR-338 (Figure 3.5). Components of the TGF-β pathway (*Tgfbr1* and *Tgfbr2*), the most potent positive regu-

lator of fibrosis, are targeted by miR-335 and the let-7 family, while components of the PDGF pathway (*Pdgfa* and *Pdgfb*), responsible for induction of HSC proliferation, are targeted by miR-29c and miR-335. Integrin and Rho-GTPase signaling, which mediate signals from the ECM, are regulated by miR-30 family members. Rho-GTPase signaling is additionally regulated by miR-335 and miR-338 (Figure 3.5). Collectively, these results indicate that the identified AF-miRNAs are negative regulators of structural, remodeling and signaling components of ECM organisation.

3.4 AF-miRNAs are downregulated and fibrosis-associated genes are upregulated in murine HCC

sRNA-seq performed on mouse SRF-VP16-driven HCC samples shows that the AF-miRNAs are significantly downregulated in mHCC (Figure 3.6A, B).

This study characterizes miRNA targeting of a subset of ECM-related genes, which covers structural (*Col1a1, Col1a2, Col4a2, Col4a2, Col4a5, Col5a2* and *Lamc1*), signaling (*Pdgfa, Pdgfb* and *Tgfbr1*) and remodeling (*Loxl2, Loxl4, Adamts14, Adamts15* and *Tpm1*) components of the ECM (Figure 3.5).

While AF-miRNAs are downregulated, all aforementioned genes, except *Col5a2*, show significant upregulation in the mouse *SRF-VP16*^{iHep} model (Figure 3.6C, D, E and F). *Col5a2* shows clear, but statistically not significant (p_{adj}-value 0.058) upregulation.

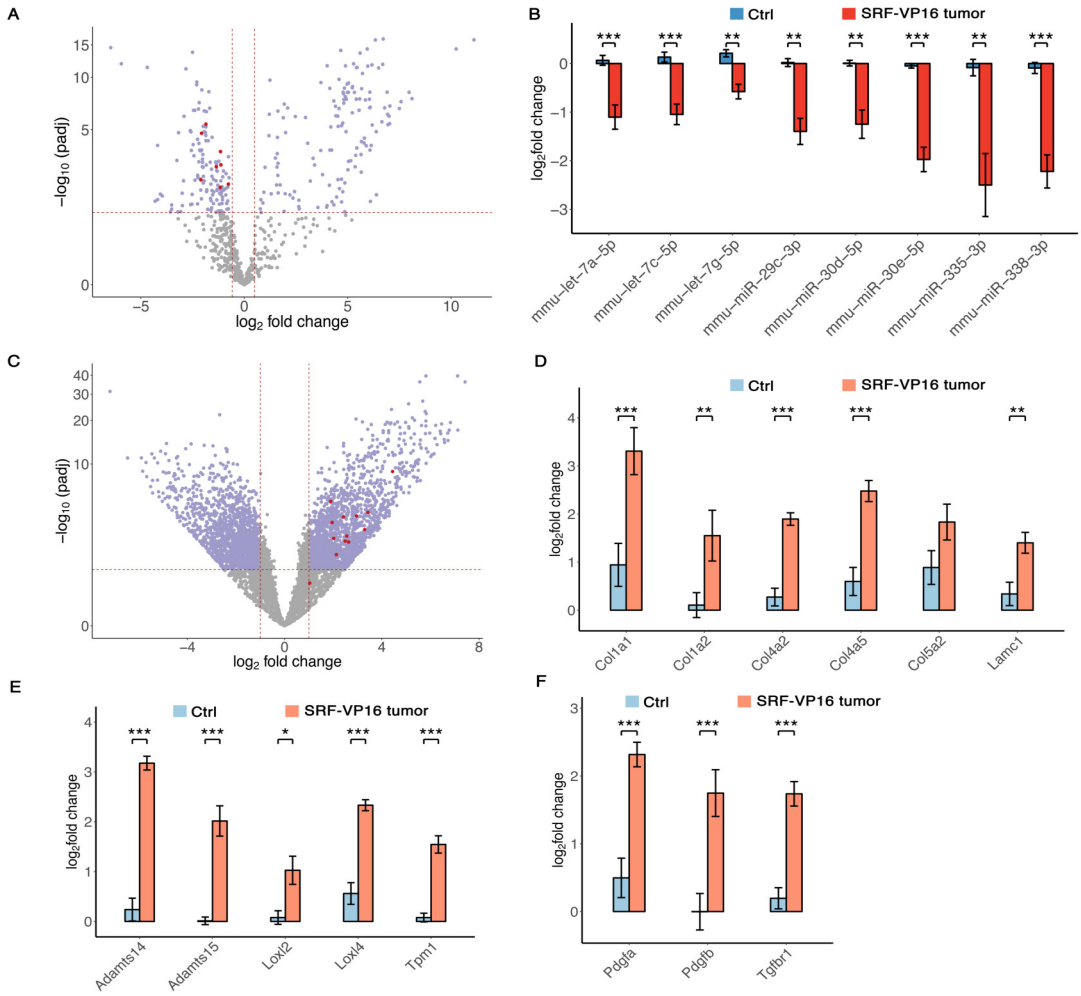

Figure 3.6: AF-miRNAs are downregulated and fibrosis-associated genes are upregulated in murine HCC. (A) Volcano plot of miRNAs identified in sRNA-seq. AF-miRNAs are depicted in red, significantly dysregulated miRNAs (threshold 1.5-fold) in violet. (B) sRNA-seq-derived, normalised read counts ($\log_2$-transformed) of AF-miRNAs in control and tumor samples of *SRF-VP16*^{iHep} mice. (C) Volcano plot of genes identified in RNA-seq. Fibrosis-related genes characterized in this study are shown in red, significantly dysregulated genes (threshold 2-fold) in violet. (D-F) Normalized read count ($\log_2$-transformed) of fibrosis-associated, structural (D), remodeling (E) and signaling (F) genes of the ECM in control and tumor samples of *SRF-VP16*^{iHep} mice. All samples are normalized to a randomly chosen control sample. Data are shown as mean and standard error of the mean. p_{adj}-value ≤ 0.05, ** p_{adj}-value ≤ 0.01, *** p_{adj}-value ≤ 0.001.

3.5 AF-miRNAs are downregulated and fibrosis-associated genes are upregulated in the pHSC *in vitro* culture fibrosis model and *in vivo* CCl_4 murine fibrosis model

To test the generality of the miRNA:mRNA network, the expression of AF-miRNAs and fibrosis-associated target genes was examined in two fibrosis models, murine pHSC *in vitro* culture and the carbon tetrachloride (CCl_4) *in vivo* mouse model.

pHSC activation, evident by loss of retinoid droplets and increased ECM production, occurs when cells are plated on standard plastic dishes[14,73]. To ensure activation of pHSCs, cells were maintained in culture for seven days (Figure 3.7). Comparing day two and day seven of cell culturing, clearly indicates changes of pHSC morphology and decrease in presence of retionid droplets. The CCl_4 *in vivo* model was generated by prolonged administration of CCl_4, which leads to hepatic fibrosis development[68].

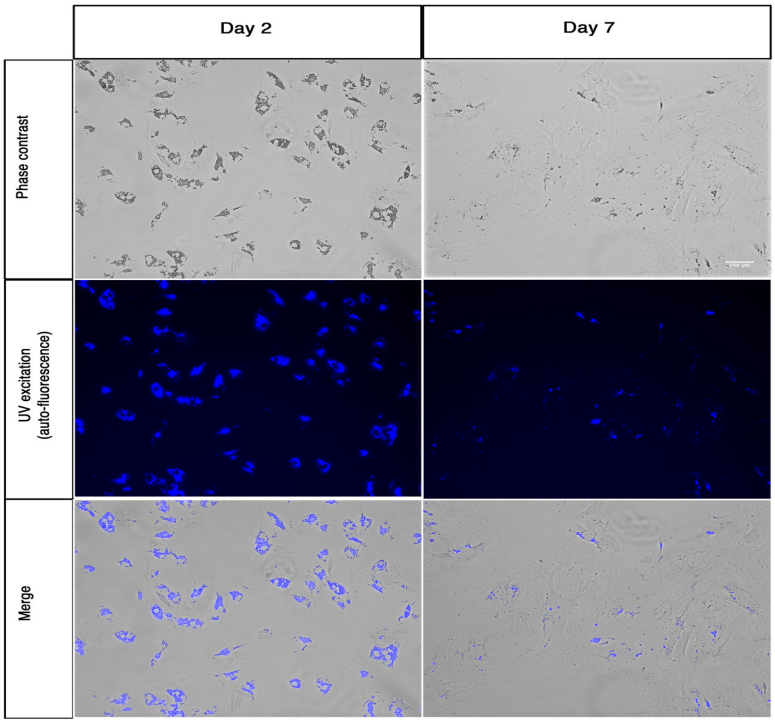

Figure 3.7: Primary murine hepatic stellate cells (pHSC) are spontaneously activated as a result of prolonged (seven days) growth in uncoated, standard plastic cell culture dishes. Upper row shows changes in morphology of pHSCs upon culturing, as visualized by phase contrast microscopy. Middle row shows auto-fluorescence of retinoid droplets in cytoplasm of pHSCs upon UV excitation. Auto-fluorescence of retinoid droplets is diminished at day 7 of pHSC culturing, indicating activation of pHSCs. Bottom row shows merge of upper and middle rows.

Using qPCR, the expressions of AF-miRNAs and ECM-related miRNA-targeted genes were compared between freshly isolated (inactive) pHSCs and pHSCs activated by prolonged culturing. In addition, the expressions of AF-miRNAs and ECM-related miRNA-targeted genes was measured in CCl_4-treated samples and their corresponding controls.

All AF-miRNAs were found downregulated in both fibrosis models (Figures 3.8A, 3.9A), while all measured structural ECM fibrosis-associated target genes were upregulated in both models (Figures 3.8B, 3.9B). Similarly, all examined remodeling and signaling ECM components were upregulated, except *Tgfbr1* and *Adamts*, which were upregulated in CCl_4 model but downregulated in the pHSC model (Figures 3.8C, D, 3.9C, D).

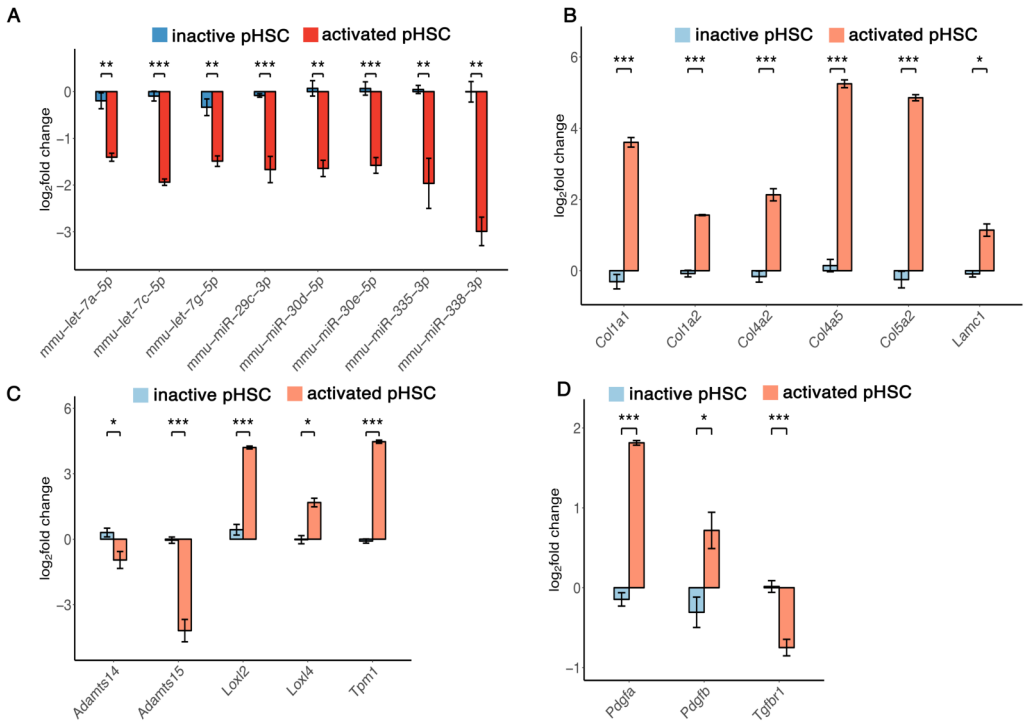

Figure 3.8: AF-miRNAs are downregulated and fibrosis-associated genes are upregulated in the pHSC *in vitro* culture fibrosis model. (A) Relative expression of mature miRNAs in inactive (freshly isolated) and activated (prolonged *in vitro* culturing) pHSCs. (B-D) Relative expression of fibrosis-associated, structural (B), remodeling (C) and signaling (D) genes of the ECM in inactive and activated pHSCs. All samples are normalized to a randomly chosen control sample. Data are shown as mean and standard error of the mean. p_{adj}-value ≤ 0.05, ** p_{adj}-value ≤ 0.01, *** p_{adj}-value ≤ 0.001.

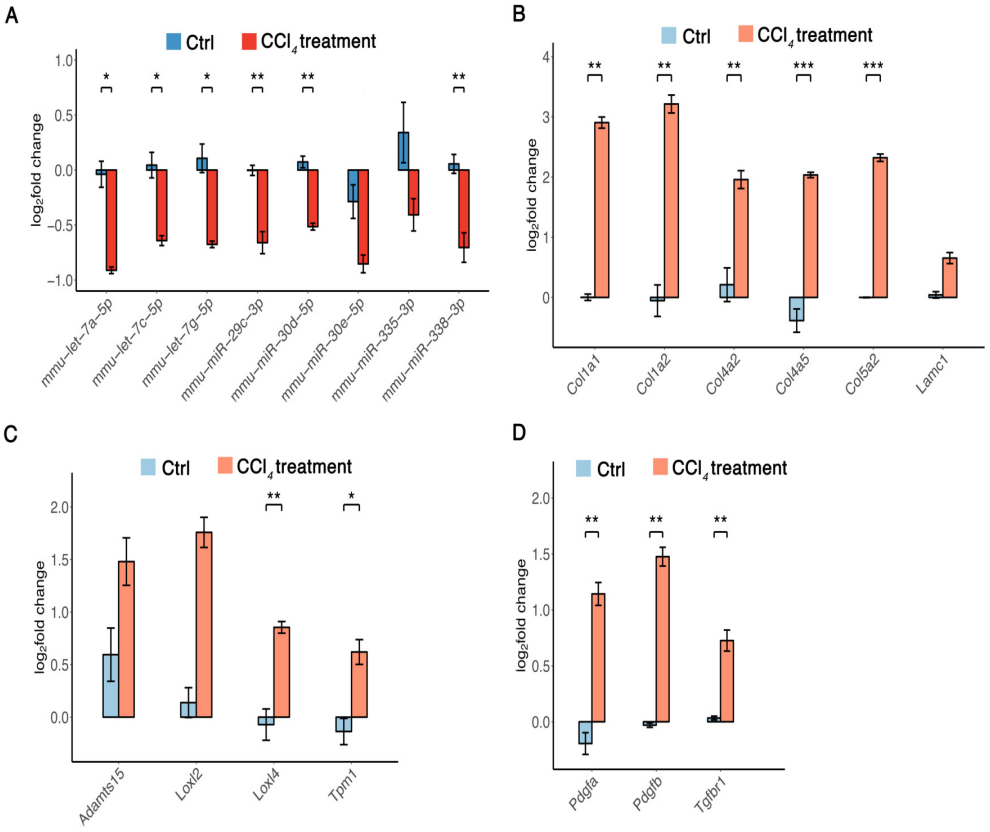

Figure 3.9: AF-miRNAs are downregulated and fibrosis-associated genes are upregulated in the *in vivo* CCl_4 murine fibrosis model. (A) Relative expression of mature miRNAs in samples of CCl_4-treated mice in comparison to controls (mineral oil treatment). (B-D) Relative expression of fibrosis-associated, structural (B), remodeling (C) and signaling (D) genes of the ECM in CCl_4-treated and control mice. All samples are normalized to a randomly chosen control sample. Data are shown as mean and standard error of the mean. * p-value ≤ 0.05, ** p-value ≤ 0.01, *** p-value ≤ 0.001.

3.6 AF-miRNAs target structural, signaling and remodeling components of the ECM

To experimentally validate the predicted AF-miRNA targeting of fibrosis-related genes, luciferase assays and qPCR analyses were performed (Table 2.16). To modulate miRNA expression, two strategies were employed. In the first approach, miRNA mimics were used to overexpress miR-29c, miR-338, let-7a, let-7c and let-7g. In the second approach, let-7a, let-7c and let-7g expression was inhibited by overexpressing Lin28a, which downregulates the entire let-7 family[106].

To perform luciferase assays, 3'-UTR regions of *Col1a1*, *Pdgfa*, *Tgfbr1* and *Adamts15* genes were cloned downstream of the luciferase gene and luciferase expression was assayed in NIH/3T3 cells upon miRNA mimic transfection. For validation, the miRNA binding sites were mutated and likewise the mutant 3'-UTR constructs were assayed upon miRNA mimic transfection.

The luciferase reporter containing wild-type 3'-UTR of *Col1a1* showed significant downregulation upon let-7a, let-7c, let-7g or miR-29c mimic expression compared to scrambled mimic, while the mutant 3'-UTR construct retained comparable levels of expression upon specific miRNA and scrambled mimic transfection (Figure 3.10A).

A similar downregulation of *Pdgfa* and *Tgfbr1* constructs was observed upon miR-29c and let-7a, let-7c and let-7g mimics overexpression, respectively (Figure 3.10B, C). The *Adamts15* luciferase construct showed a significant downregulation upon miR-338, let-7a, let-7c, let-7g and miR-29c mimic transfection (Figure 3.10D). In contrast, all the constructs with mutated miRNA binding sites expressed high luciferase mRNA levels upon specific miRNA and scrambled mimic transfection (Figure 3.10B, C, D).

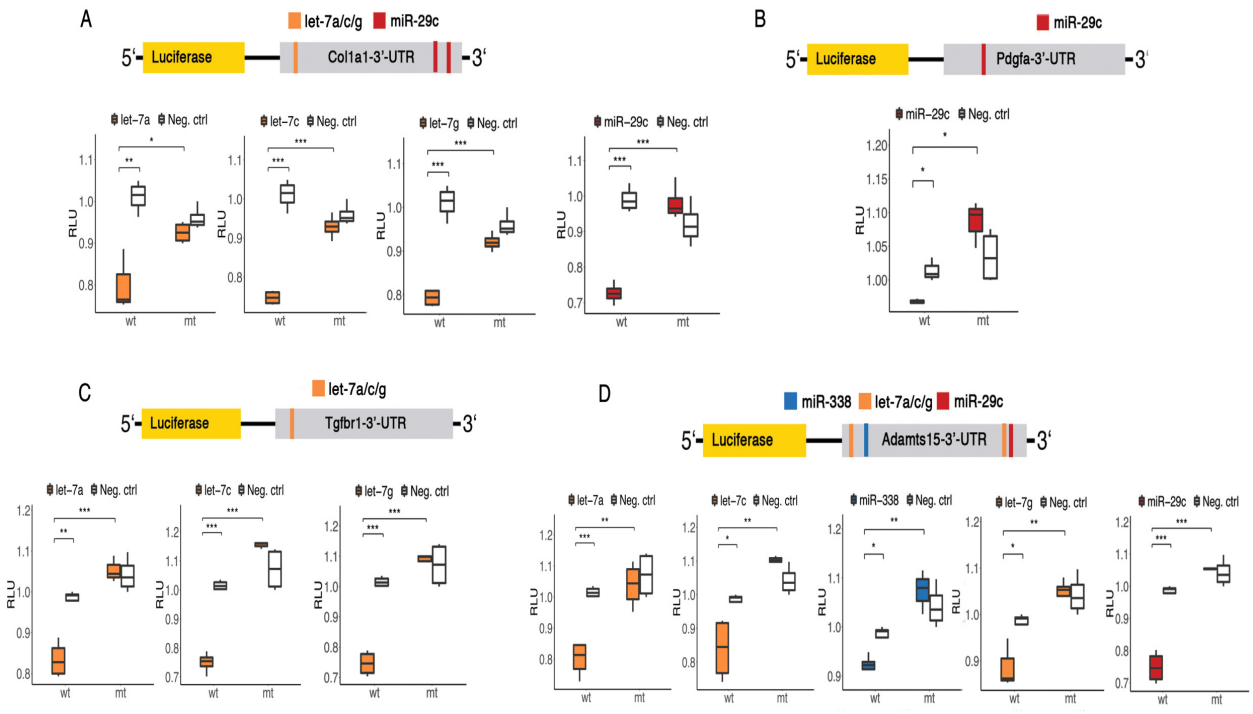

Figure 3.10: AF-miRNAs target structural, signaling and remodeling components of the ECM. (A-D) Activities of wild-type and mutant (mutated miRNA site) 3'-UTR-luciferase constructs derived from: (A) *Col1a1* in NIH/3T3 cells transfected with let-7a, let-7c, let-7g, miR-29c and scrambled miRNA mimic (Neg. ctrl), (B) *Pdgfa* in NIH/3T3 cells transfected with miR-29c and scrambled miRNA mimic, (C) *Tgfbr1* in NIH/3T3 cells transfected with let-7a, let-7c, let-7g and scrambled miRNA mimic, (D) *Adamts15* in NIH/3T3 cells transfected with miR-338, let-7a, let-7c, let-7g, miR-29c and scrambled miRNA mimic. Let-7a-, let-7c-, let-7g-, miR-29c- and miR-338-transfected samples are colored in the plots according to the luciferase construct schematic. Samples transfected with scrambled miRNA mimic are shown in white. Data are shown as median, first and third quartile ("hinges") and 95% confidence interval of median ("notches"). * p-value $\leq$ 0.05, ** p-value $\leq$ 0.01, *** p-value $\leq$ 0.001.

In qPCR analysis, the effect of the let-7 family on target gene expression was assessed in Lin28a-overexpressing NIH/3T3 cells (Figure 3.11A-C). Lin28a overexpression relative to control was validated and its inhibitory effect on let-7a, let-7c and let-7g was confirmed (Figure 3.11A, B) prior to quantification of the expression of miRNA target genes. Lin28a-mediated inhibition of the let-7 family resulted in signifi-

cant upregulation of most collagens. While *Tgfbr1* and *Loxl4* showed a similar trend of upregulation, *Col4a2* and *Adamts14* appeared to be weakly downregulated (Figure 3.11C).

Additionally, the effect of let-7c, let-7g and miR-29c upregulation (Figure 3.11D, F, H) on target gene expression (Figure 3.11E, G, I) upon miRNA mimic or scrambled mimic transfection was investigated. All three miRNAs caused specific downregulation of their target genes. While miR-29c and let-7c mediate significant inhibition of all their target genes with the exception of Col4a2 for let-7c (Figure 3.11E, I), the inhibitory effect of let-7g is more subtle (Figure 3.11G).

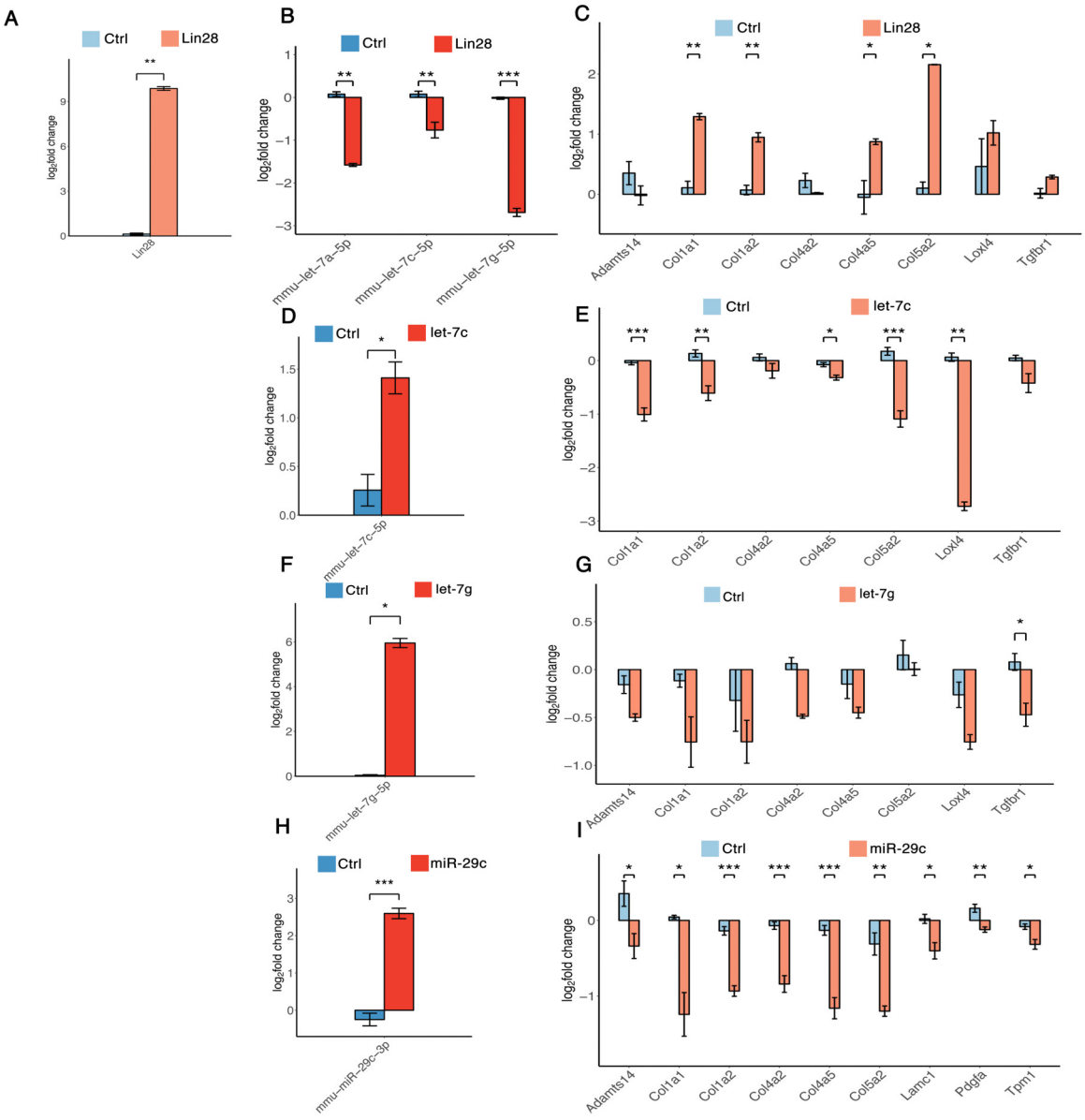

Figure 3.11: AF-miRNAs target structural, signaling and remodeling components of the ECM. (A-C) Relative expression of: (A) *Lin28a*, (B) let-7a, let-7c and let-7g and (C) putative let-7 target genes associated with fibrosis in stable Lin28a-overexpressing NIH/3T3 cells. (D-E) Relative expression of: (D) let-7c and (E) putative let-7 target genes associated with fibrosis in NIH/3T3 cells transfected with let-7c mimics. (F-G) Relative expression of: (D) let-7g and (E) putative let-7 target genes associated with fibrosis in NIH/3T3 cells transfected with let-7g mimics. (H-I) Relative expression of: (H) miR-29c and (I) putative miR-29c target genes associated with fibrosis in NIH/3T3 cells transfected with miR-29c mimics. Data are shown as mean and standard error of the mean. * p-value $\leq$ 0.05, ** p-value $\leq$ 0.01, *** p-value $\leq$ 0.001.

3.7 AF-miRNAs are downregulated and fibrosis-associated genes are upregulated in a subset of human HCCs

Expression analysis of the AF-miRNAs in the hHCC dataset showed that the same miRNAs are downregulated in at least 10% of TCGA cases with the exception of hsa-miR-30d-5p, which is downregulated in only 3.2% of cases (Figure 3.12A).

The miRNAs hsa-let-7c-5p and hsa-miR-29c-3p show the highest frequency of downregulation, 71% and 66%, respectively. Hsa-miR-335-3p and hsa-miR-338-3p are downregulated in smaller number of cases, 30.1% and 10%. Although hsa-miR-30d-5p is downregulated in a very small fraction of tumors, hsa-miR-30e-5p, which shares the same targets as miR-30d, is downregulated in 38% of cases (Figure 3.12A).

In the human TCGA cohort, most of the fibrosis-associated genes show upregulation ($\geq$ 2-fold) in the majority of tumors with the exception of *TGFBR1* and *ADAMTS15*, which show upregulation in 11.8% and 7.5% of cases, respectively (Figure 3.12B).

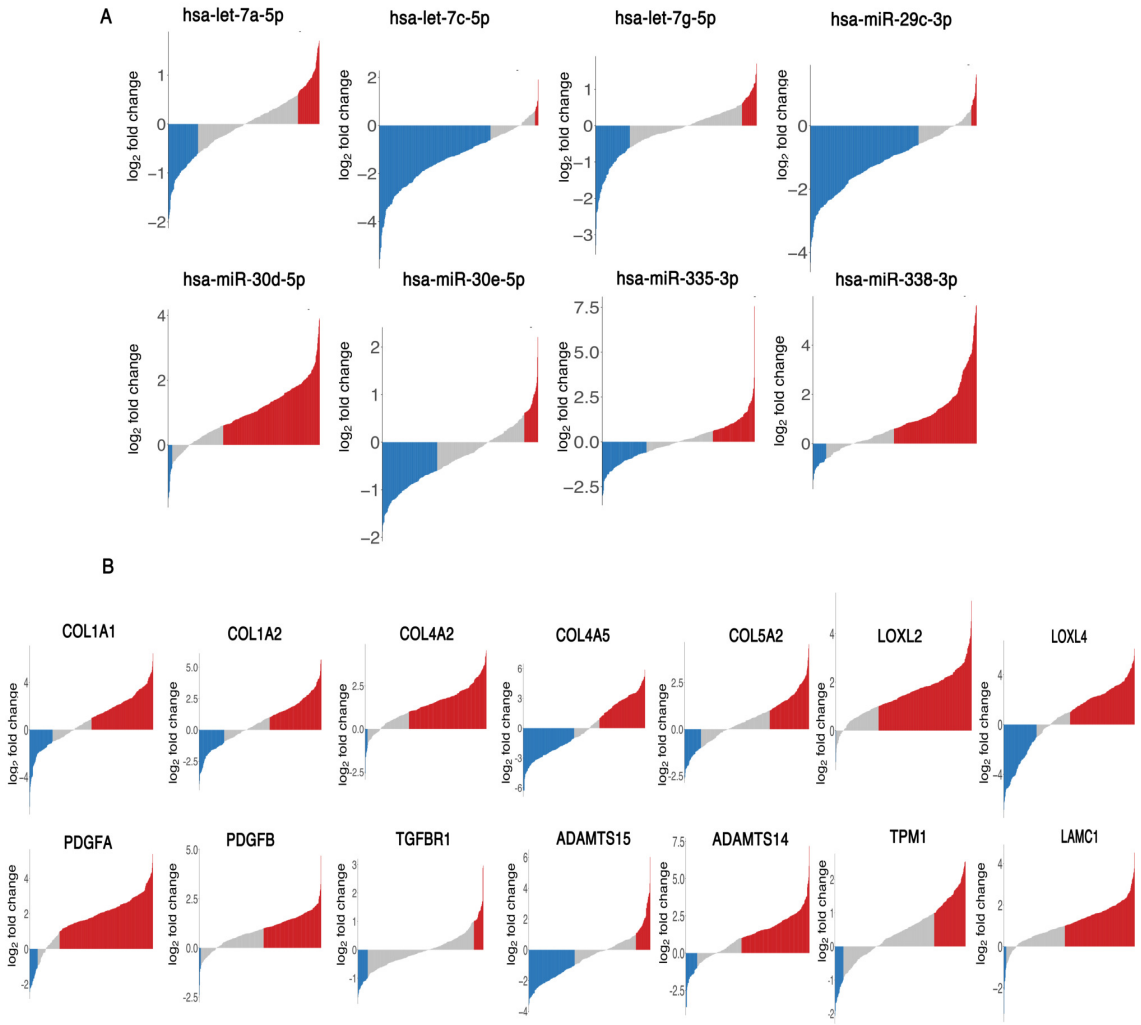

Figure 3.12: AF-miRNAs are downregulated and fibrosis-associated genes are upregulated in a subset of human HCCs. (A) Log_2-fold changes of AF-miRNAs in individual tumors (compared to the mean control value) of the human TCGA dataset. Bars in waterfall plot: blue (miRNA downregulation $\geq$ 1.5-fold), red (miRNA upregulation $\geq$ 1.5-fold). (B) Log_2-fold changes of fibrosis-associated genes in individual tumors (compared to the mean control value) of the human TCGA dataset. Bars in waterfall plot: blue (mRNA downregulation $\geq$ 2-fold), red (mRNA upregulation $\geq$ 2-fold).

3.8 The miRNA:mRNA pairs show different degrees of association in human fibrosis-facilitated carcinomas

To examine the recurrence of expression association of the AF-miRNAs and their target mRNAs in a wider range of human carcinomas, the multivariate linear regression approach[103] was implemented.

The multivariate model assesses mRNA expression regulation taking into consideration: miRNA and mRNA expression, changes in DNA copy number (CNV) and promoter methylation status of the respective gene. CNV and methylation data are used to assess the influence of miRNA-unrelated gene regulation. The aim was to assess, using the multivariate model, if AF-miRNAs considerably contribute to fibrosis-associated gene regulation *in vivo* in different fibrosis-facilitated carcinomas.

The model was implemented across HCC, BRCA, LUAD and LUSC, i.e. carcinoma types exhibiting fibrotic tumor microenvironments which supports tumor progression[10].

First, the recurrence of expression association of the miRNAs and their target mRNAs in individual carcinoma types was examined. In hHCC, most ECM-related genes are consistently targeted by let-7g and miR-29c, while Rho GTPase-related genes are targeted by miR-30e (Figure 3.13). Similarly, let-7g and miR-29c consistently inhibit the majority of ECM-related genes in BRCA and LUAD samples. However, in these carcinomas, miR-335 and miR-338 additionally contribute to targeting of ECM genes, while Rho GTPase-related genes are primarily modulated by miR-30d and miR-338 (Figures 3.14, 3.15, 3.16). Furthermore, miR-335 regulates ECM- and integrin-related genes in LUAD and LUSC (Figure 3.15, 3.16).

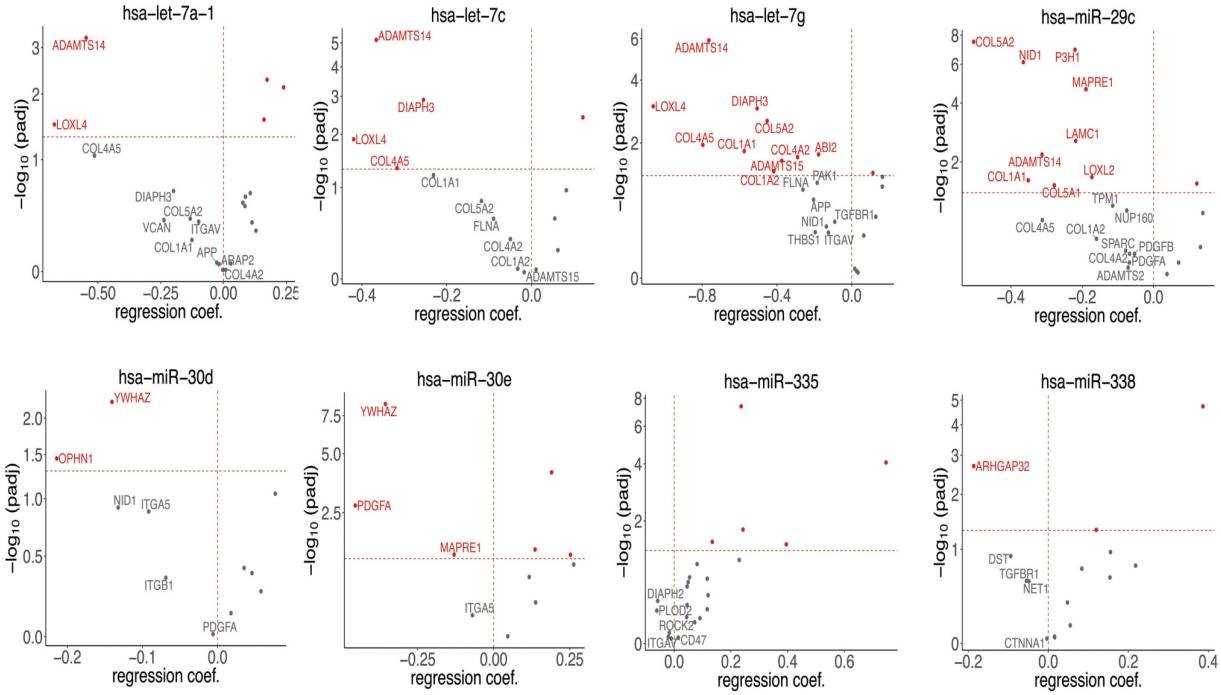

Figure 3.13: The miRNA:mRNA pairs show different degrees of association in human HCC carcinomas. Each plot shows recurrence of expression association of the particular miRNA and its ECM-related target mRNA in human HCCs defined by regression coefficient and corresponding p_{adj} value. The relationship of miRNA to individual mRNA was evaluated using a multivariate linear model which factors in: miRNA and mRNA expression, changes in DNA copy number (CNV) and promoter methylation status of the protein-coding genes. CNV and methylation data were used to assess the influence of miRNA-unrelated gene regulation. Target genes which have a p_{adj} value of miRNA:mRNA association assessment ≤ 0.05 are represented in red. Gene names are displayed if their regression coefficient has negative value. * p-value ≤ 0.05, **p-value ≤ 0.01, *** p-value ≤ 0.001.

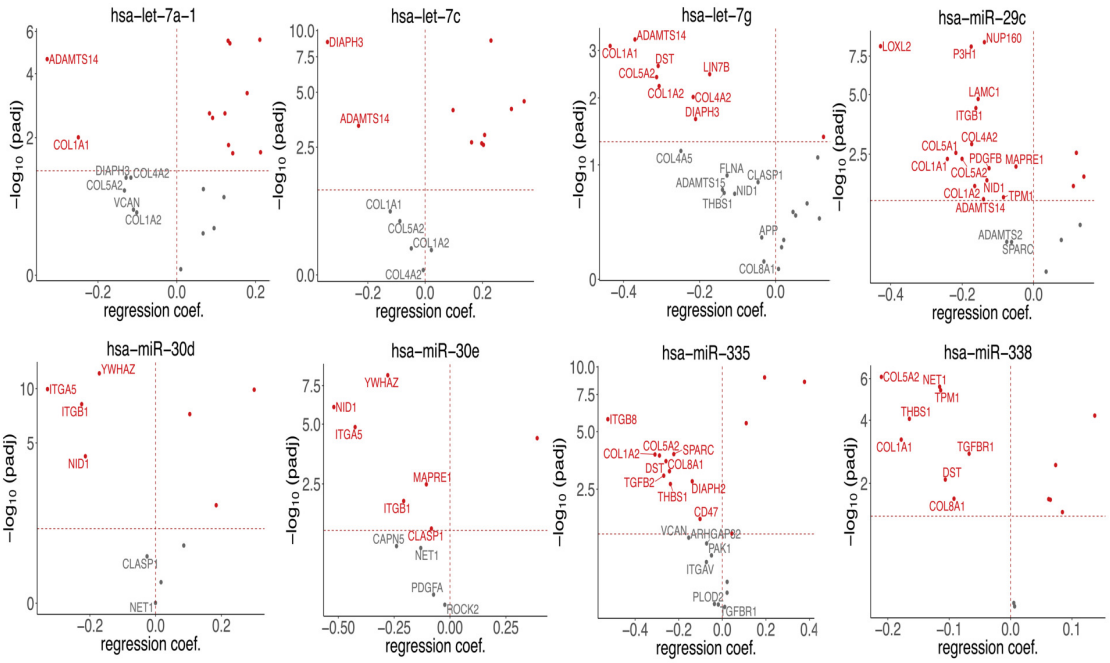

Figure 3.14: miRNA:mRNA pairs in human invasive breast carcinoma show different degrees of association in comparison to other fibrosis-facilitated carcinomas. Each plot shows recurrence of expression association of the particular miRNA and its ECM-related target mRNA in human invasive breast carcinomas defined by regression coefficient and corresponding p_{adj} value. Target genes having a p_{adj} value of miRNA:mRNA association assessment of ≤ 0.05 are represented in red. Gene names are displayed if their regression coefficient has negative value. * p-value ≤ 0.05, **p-value ≤ 0.01, *** p-value ≤ 0.001.

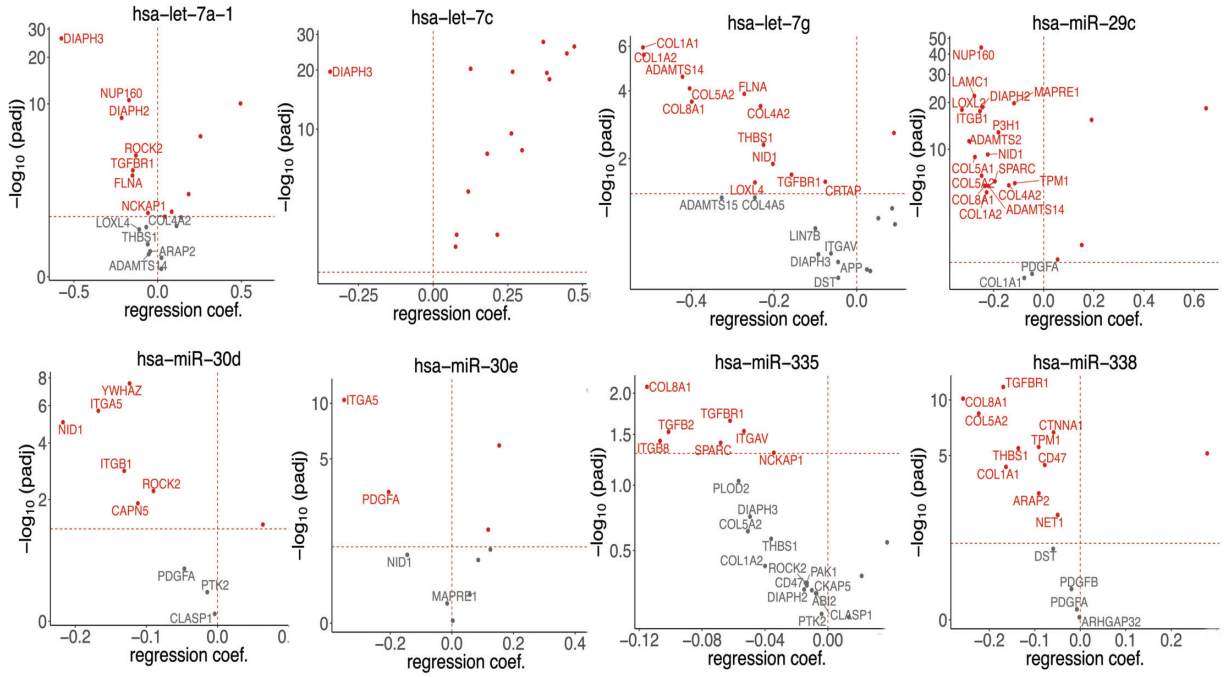

Figure 3.15: miRNA:mRNA pairs in human lung adenocarcinoma show different degrees of association in comparison to other fibrosis-facilitated carcinomas. Plots show recurrence of expression association of the particular miRNA and its target mRNA in human lung adenocarcinoma defined by regression coefficient and corresponding p_{adj} value. Target genes having a p_{adj} value of miRNA:mRNA association assessment of ≤ 0.05 are represented in red. Gene names are displayed if their regression coefficient has negative value. * p-value ≤ 0.05, **p-value ≤ 0.01, *** p-value ≤ 0.001.

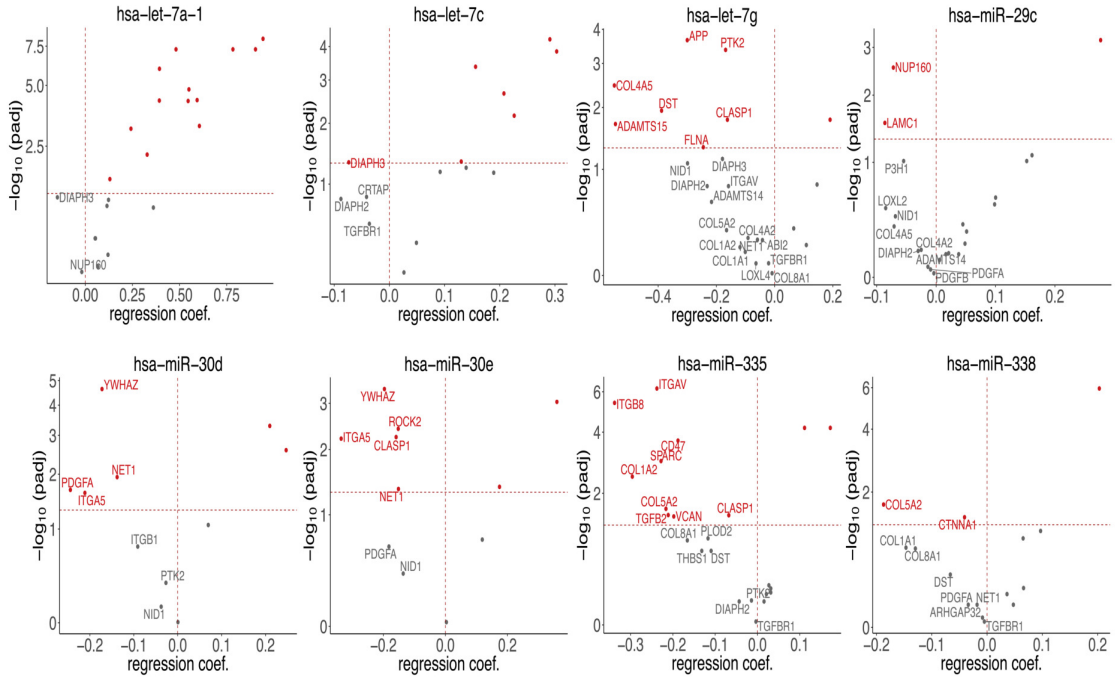

Figure 3.16: miRNA:mRNA pairs in human lung squamous cell carcinoma show different degrees of association in comparison to other fibrosis-facilitated carcinomas. Plots show recurrence of expression association of the particular miRNA and its target mRNA in human lung squamous cell carcinoma defined by regression coefficient and corresponding p_{adj} value. Target genes having a p_{adj} value of miRNA:mRNA association assessment of ≤ 0.05 are represented in red. Gene names are displayed if their regression coefficient has negative value. * p-value ≤ 0.05, **p-value ≤ 0.01, *** p-value ≤ 0.001.

Second, the association recurrence across all examined carcinomas, defined by the association recurrence (REC) score was analysed. A negative REC score of miRNA:mRNA pairs across fibrosis-facilitated carcinomas indicates that miR-29c, let-7g, let-7a and miR-335 consistently regulate different collagens (Figure 3.17). *Adamts14* is regulated by miR-29c, let7a and let-7g, while *Adamts15* is targeted by let-7g and let-7c. *Loxl2* and *Loxl4* are modulated by miR-29c and let-7c and let-7a, respectively. *Tgfbr1* is primarily regulated by miR-338, let-7a and let-7c. The linear regression analysis, which assessed the extent of miRNA, DNA methylation and CNV influence on gene expression, showed that identified miRNAs considerably contribute to the expression regulation of the fibrosis-associated genes. Together, the AF-miRNAs commonly regulate ECM-related genes in human fibrosis-driven carcinomas, albeit to different extents in different carcinoma types.

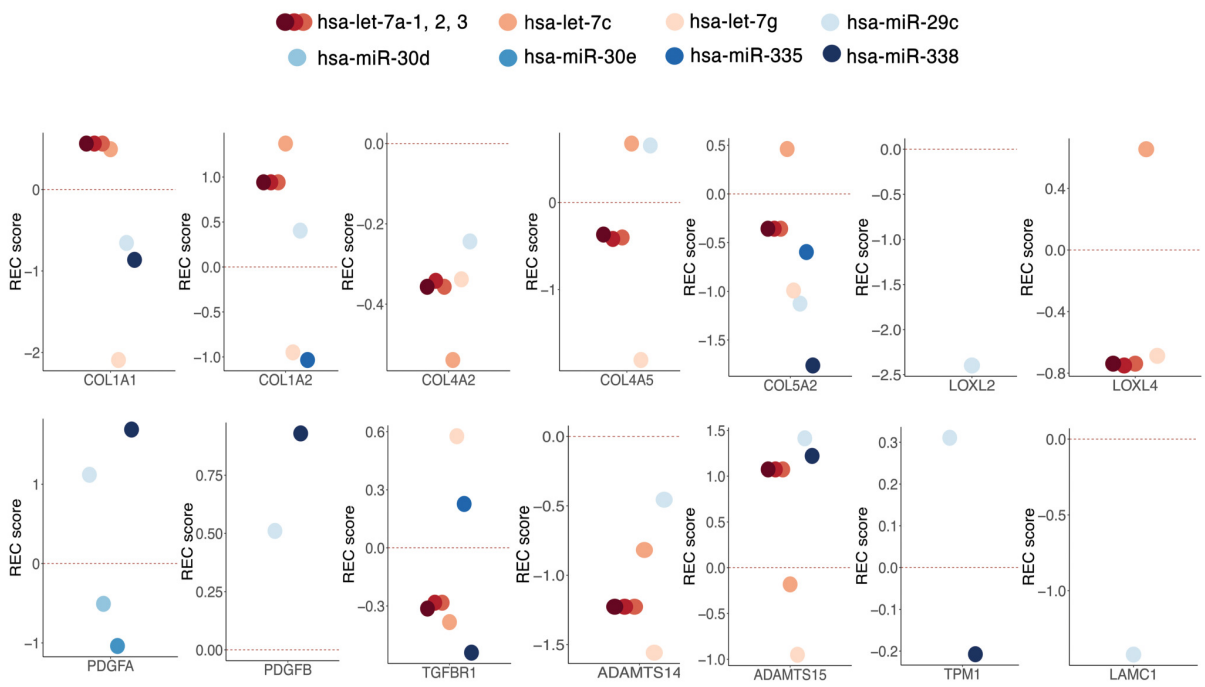

Figure 3.17: The miRNA:mRNA pairs show different degrees of association in human fibrosis-facilitated carcinomas. Inferred association recurrence (REC) scores of miRNA:mRNA pairs depicted in Figure 3.5 in fibrosis-facilitated cancers (HCC, BRCA, LUAD and LUSC). The sign of the REC score indicates the nature of the association, while its magnitude captures the recurrence consistency.

3.9 AF-miRNAs are downregulated at the pri-miRNA level suggesting transcriptional regulation of pri-miRNA-encoding genes during fibrosis

To identify the mechanism of AF-miRNA regulation during pHSC activation the expression of precursor molecules of AF-miRNAs was examined. As miRNAs can be regulated at different stages of biogenesis, pri-miRNAs and pre-miRNAs of AF-miRNAs were quantified using qPCR in inactive and activated pHSCs, as well as in the *in vivo* CCl_4 fibrosis mouse model.

The results of the analysis showed that downregulation of AF-miRNAs occurs at the pri-miRNA level in both fibrosis models, suggesting transcriptional regulation of pri-miRNA-encoding genes (Figure 3.18A, C). While all pri-miRNAs are significantly downregulated in the pHSC model, pri-miRNA downregulation in the CCl_4 model is clear, albeit less pronounced. In agreement, pre-miRNAs in the pHSC model are downregulated, as well (Figure 3.18B).

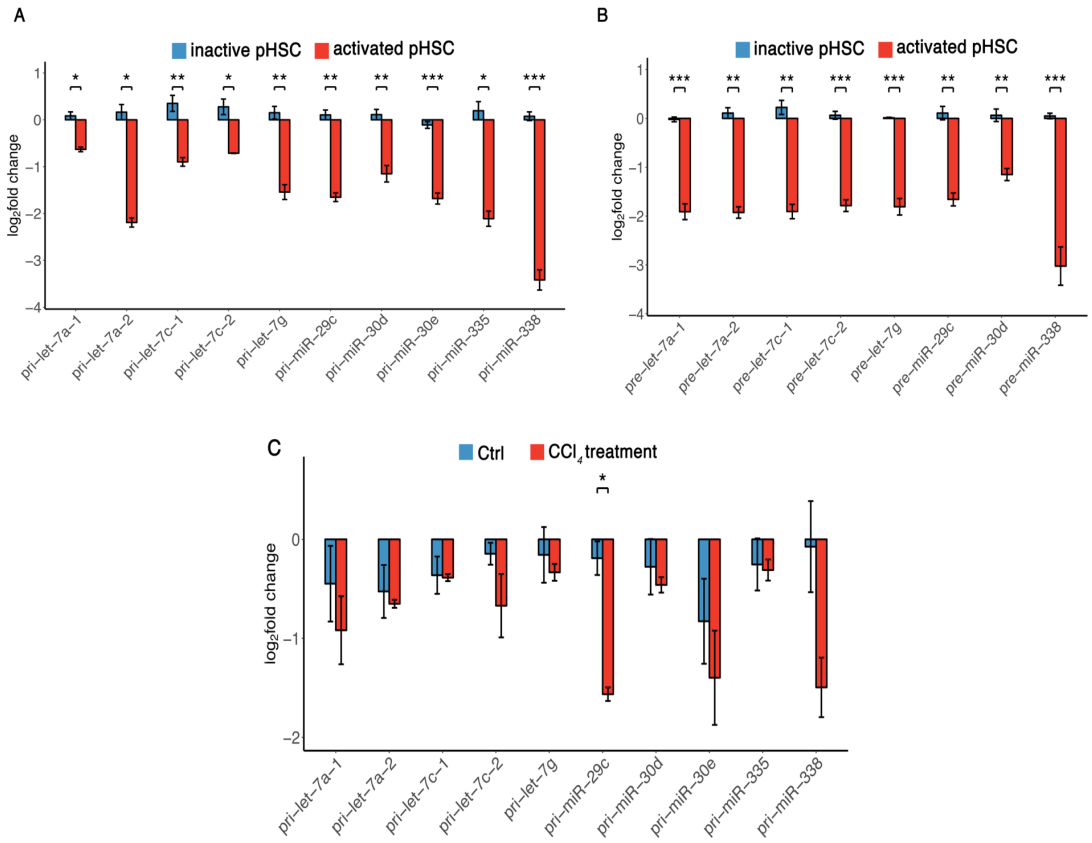

Figure 3.18: AF-miRNAs are downregulated at the pri-miRNA level suggesting transcriptional regulation of pri-miRNA-encoding genes. (A) Relative expression of AF-pri-miRNAs in the pHSC *in vitro* culture model. (B) Relative expression of pre-miRNA in inactive and activated pHSCs. (C) Relative expression of pri-miRNAs in samples of CCl_4-treated mice in comparison to controls (mineral oil treatment). Data are shown as mean and standard error of the mean. * p-value ≤ 0.05, ** p-value ≤ 0.01, *** p-value ≤ 0.001.

3.10 Transcription factors Pparγ and Egr1 regulate expression of AF-miRNAs

To study how AF-miRNAs are regulated at the transcriptional level, two major transcriptional regulation mechanisms were examined: binding of potential transcription factors to the promoters of relevant miRNA-encoding genes and CpG methylation changes in promoters of miRNA-encoding genes upon HSC activation.

First, to identify miRNA promoters publicly available GRO-seq datasets were utilised. As pri-miRNAs are rapidly cleaved to pre-miRNAs, mapping miRNA TSS using conventional TSS mapping approaches is highly challenging[97]. GRO-seq is a technique used to quantifying nascent transcripts. Briefly, transcription is halted, nuclei are isolated and transcription briefly restarted using labelled nucleotides resulting in labelled RNA molecules. Labelled RNA molecules are precipitated and sequenced[98]. GRO-seq data show sharp peaks around TSSs in both the sense and antisense directions and continuous signal of lower intensity throughout the entire transcript allowing to map TSSs of very transient transcripts[98]. Nine mouse and six human GRO-seq datasets, which were retrieved from the GEO database, were used in miRNA TSS analysis. The source tissues/cells used to generate the retrieved datasets were primary fibroblasts, embryonic stem cells, embryonic fibroblasts and liver tissue samples. Analysis of all datasets yielded the consensus TSSs of miRNA-encoding genes which are present in the large majority of the datasets.

Locations of the identified miRNA promoters display high conservation in human and mouse genomes (Figure 3.19A, B), indicating conservation of AF-miRNA regulation between mice and humans. To examine if a majority of AF-miRNAs is regulated by common transcription factors, the FIMO tool of the MEME suite[101] was used to predict transcription factor binding to the mouse and human promoters of AF-miRNA-encoding genes.

To select final candidates of transcription factors, three filtering criteria were applied: (i) DNA binding prediction with $\leq$ 10% FDR, (ii) binding to the majority of mouse and human miRNA promoters and (iii) differential expression ($p_{adj} \leq 0.05$) in inactive versus activated hepatic and pancreatic stellate cells (GEO datasets).

This approach identified Pparγ and Egr1 as potential transcription regulators of mouse and human AF-miRNA. Pparγ is predicted to regulate mouse and human let-7a, let-7g, miR-338, miR-29c, miR-30e and miR-30d genes while Egr1 is predicted

to regulate let-7a, let-7g, miR-335, miR-30e and miR-30d. Pparγ was additionally identified as potential regulator of mouse let-7c and human miR-335. Likewise, Egr1 was identified as a regulator of mouse let-7c as well as human miR-29c and miR-338 (Figure 3.19A, B). Overlap of the predicted binding of the Egr1 and Pparγ transcription factors to the promoters of AF-miRNAs-encoding genes further indicates conservation of AF-miRNA regulation in mice and humans.

Next, the expressions of *Pparg* and *Egr1* were examined in the two fibrosis models using qPCR. The analysis showed that transcription factor *Pparg* is significantly downregulated upon pHSC activation in both fibrosis models (Figure 3.19C). *Egr1*, however, is significantly downregulated upon pHSC activation in pHSC *in vitro* model, but upregulated in the murine CCl_4 fibrosis model (Figure 3.19D).

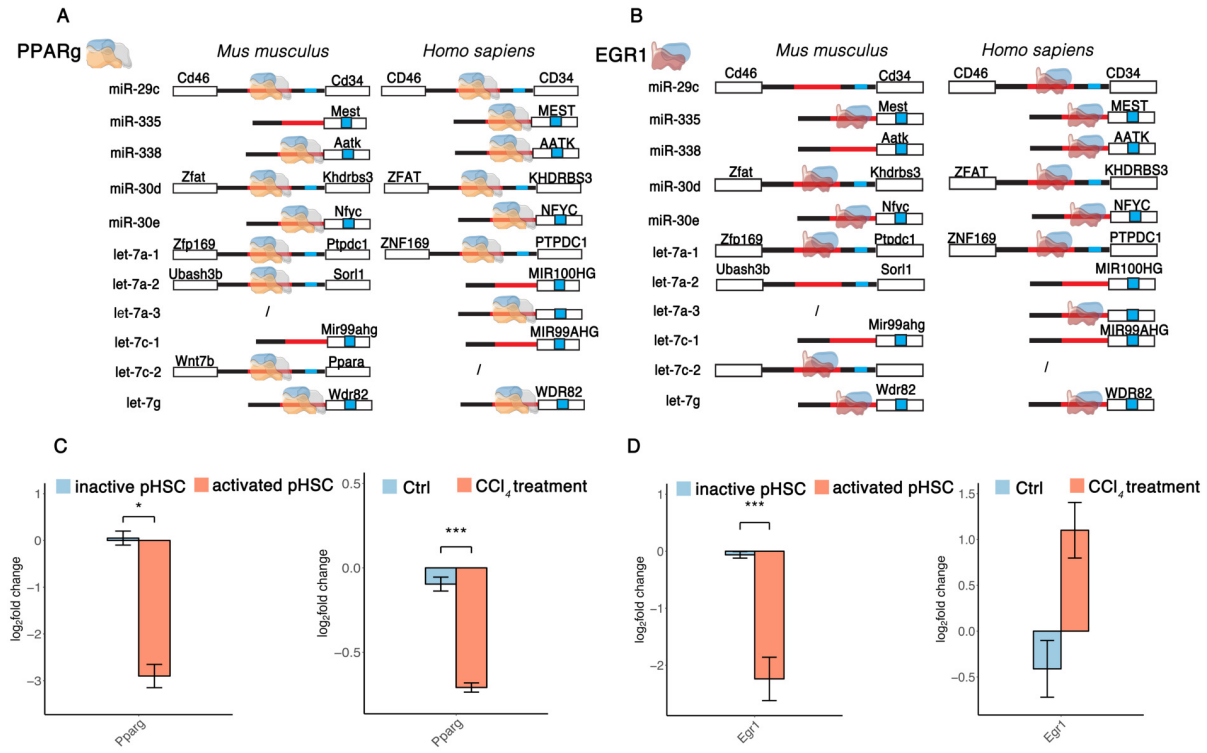

Figure 3.19: Transcription factors Pparγ and Egr1 are predicted to regulate expression of AF-miRNAs. (A) Schematic representation of predicted Pparγ binding to *M. musculus* and *H. sapiens* miRNA-encoding gene promoters. MiRNA-encoding gene promoters are shown in red, whereas mature miRNAs are indicated in blue. MiRNAs located in exons or introns of protein-coding genes share the promoter of the respective protein-coding genes. In the cases of miRNAs located in intergenic regions of the genome, the nearest neighbouring genes are shown. Pparγ is depicted on the individual miRNA-encoding gene promoter if it is predicted to bind to the respective promoter. (B) Schematic representation of predicted Egr1 binding to *M. musculus* and *H. sapiens* miRNA-encoding gene promoters. Egr1 is depicted on the individual miRNA-encoding gene promoter if it is predicted to bind to the respective promoter. (C) Relative expression of *Pparg* in the pHSC *in vitro* culture model and in the murine CCl_4 model. (D) Relative expression of *Egr1* in the pHSC *in vitro* culture model and in the murine CCl_4 model. Data are shown as mean and standard error of the mean. * p-value $\leq$ 0.05, ** p-value $\leq$ 0.01, *** p-value $\leq$ 0.001.

Next, Pparγ and Egr1 binding to the identified miRNA-encoding gene promoters was tested experimentally using ChIP. Due to the high number of cells required to perform the procedure, it was not possible to perform ChIP on pHSCs. Instead, as it was postulated based on results obtained from analysis of transcription factor expression in fibrosis models, that reduced miRNA expression is a consequence of *Pparg* and

Egr1 downregulation, a stable Pparγ-overexpressing GRX hepatic stellate cell line and stable Egr1-overexpressing NIH/3T3 cell line were generated. Pparγ-overexpressing and Egr1-overexpressing cells together with corresponding control cells were used: (i) to quantify mature miRNA expression upon transcription factor upregulation and (ii) to perform ChIP.

All AF-miRNAs predicted to be regulated by Pparγ and Egr1 show a definite trend of upregulation upon Pparγ and Egr1 overexpression (Figures 3.20A, 3.21A), respectively. MiR-335 could not be quantified due to low expression in the GRX and NIH/3T3 cell lines.

ChIP analysis showed high-affinity binding of Pparγ and Egr1 to the promoters of anti-fibrotic miRNA-encoding genes in cells overexpressing Pparγ and Egr1, respectively. Binding of both Pparγ and Egr1 was drastically lower in control GRX and NIH/3T3 cells, respectively, demonstrating that expression levels of Pparγ and Egr1 directly modulate their binding to miRNA-encoding gene promoters (Figures 3.20B, 3.21B).

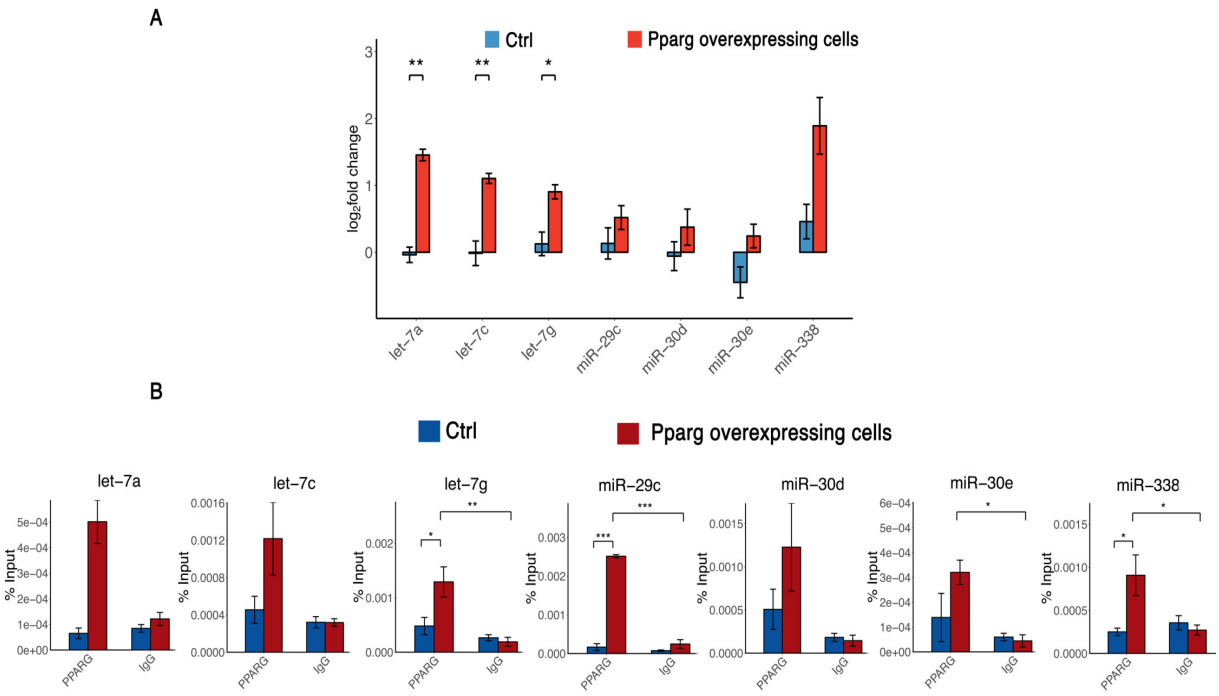

Figure 3.20: Transcription factor Pparγ regulates expression of AF-miRNAs. (A) Relative expression of AF-miRNAs in stable Pparγ-overexpressing GRX hepatic stellate cell line. (B) ChIP analysis of Pparγ binding to the promoters of AF-miRNAs in Pparγ-overexpressing GRX (red bars) and control GRX hepatic stellate cell line (blue bars). Data are shown as mean and standard error of the mean. * p-value $\leq$ 0.05, ** p-value $\leq$ 0.01, *** p-value $\leq$ 0.001.

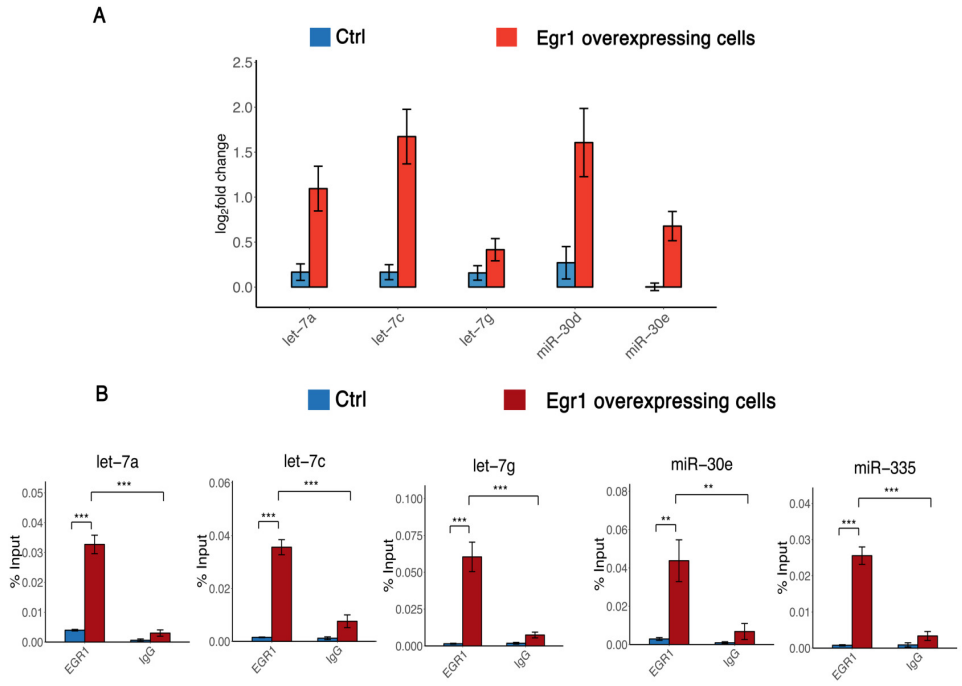

Figure 3.21: Transcription factor Egr1 regulates expression of AF-miRNAs. (A) Relative expression of AF-miRNAs in stable Egr1-overexpressing NIH/3T3 cell line. (E) ChIP analysis of Egr1 binding to the promoters of AF-miRNAs in Egr1-overexpressing NIH/3T3 (red bars) and control NIH/3T3 cell line (blue bars). Data are shown as mean and standard error of the mean. * p-value ≤ 0.05, ** p-value ≤ 0.01, *** p-value ≤ 0.001.

3.11 DNA methylation of miRNA-encoding gene promoters contributes to the regulation of AF-miRNA expression

As genome CpG methylation influences significantly gene expression, methylation changes in the promoters of miRNA-encoding genes were examined in pHSCs. Significant hypermethylation of let-7a, miR-335 and miR-338 gene promoters was observed upon pHSC activation, suggesting epigenetic mechanisms to contribute to transcriptional control of AF-miRNAs. Although not significant, let-7c-1 and miR-30e promoters were also hypermethylated upon pHSC activation (Figure 3.22).

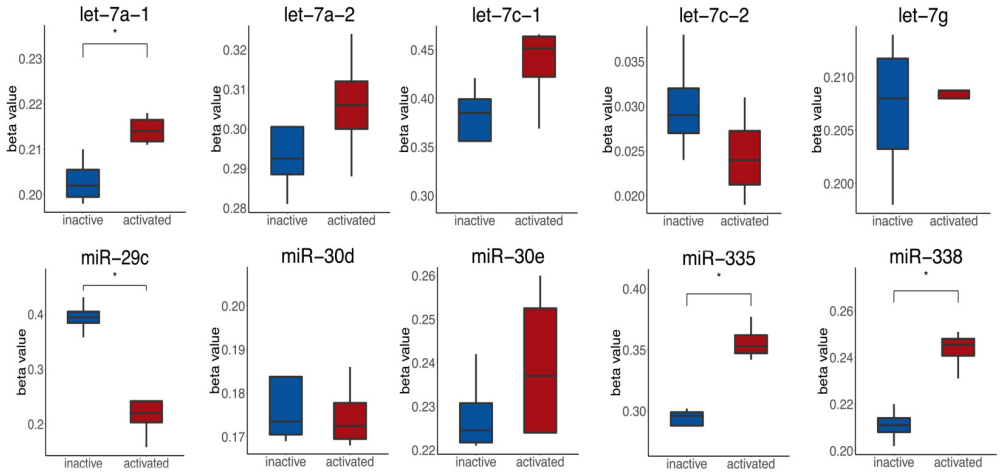

Figure 3.22: DNA methylation of miRNA-encoding gene promoters contributes to the regulation of AF-miRNAs expression. Differential CpG DNA methylation of promoters of AF-miRNA-encoding genes in inactive and activated pHSCs. Data are shown as median, first and third quartile ("hinges") and 95% confidence interval of median ("notches"). * p-value ≤ 0.05, ** p-value ≤ 0.01, *** p-value ≤ 0.001.

In conclusion, this study showed that AF-miRNAs are significantly downregulated in the *SRF-VP16*^{iHep}mHCC model, while fibrosis-associated genes are upregulated. Additionally, it was shown that AF-miRNAs are significantly downregulated upon pHSCs activation. CCl_4 treatment of mice, which induces fibrosis led to downregulation of AF-miRNAs. Furthermore, using a linear regression model, it was demonstrated that AF-miRNAs significantly contribute to the regulation of fibrosis-associated genes in human fibrosis-associated carcinomas, i.e. HCC, BRCA, LUAD and LUSC. Also, modulation of AF-miRNA expression (let-7a, let-7c, let-7g and miR-29c) downregulated expression of fibrosis-associated target genes. *In vitro* luciferase assays experimentally confirmed predicted targeting of *Col1a1* by miR-29c, let-7a, let-7c and let-7g; *Pdgfa* by miR-29c; *Tgfbr1* by let-7a, let-7c and let-7g and *Adamts15* by miR-29c, miR-338, let-7a, let-7c and let-7g. Therefore, all results outlined above support the conclusion that AF-miRNAs indeed act as anti-fibrotic miRNAs.

Based on these conclusions and the described findings of AF-miRNA regulation, the following model is postulated: upon HSC activation, *Pparγ* and *Egr1* expression decreases, consequently causing reduced transcription of anti-fibrotic pri-miRNAs. Re-

duced miRNA expression releases their inhibitory effect on fibrosis-associated target mRNAs, thereby increasing the abundance of pro-fibrotic proteins (Figure 3.23).

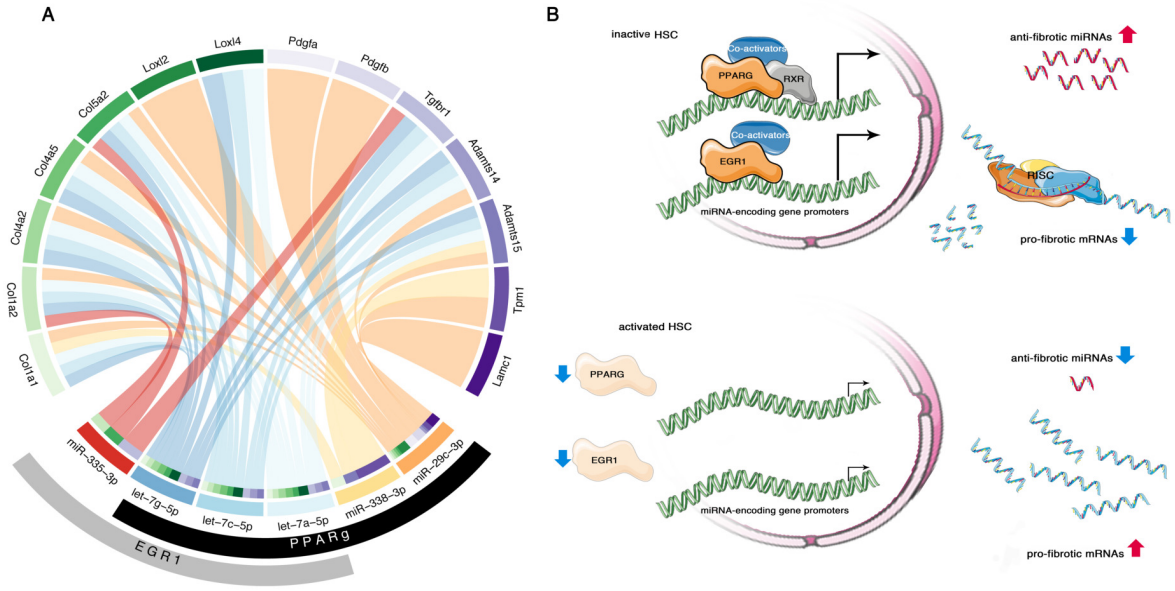

Figure 3.23: Schematic summarizing the regulation of AF-miRNAs and their target genes. (A) Circos plot summarizing the regulation of AF-miRNAs and their target genes. (B) Graphical model summarizing the regulation of AF-miRNAs. Upon HSC activation, *Pparγ* and *Egr1* expression decreases, thus leading to reduced transcription of anti-fibrotic pri-miRNAs. Reduced miRNA expression increases abundance of the fibrosis-associated target mRNAs, thereby increasing the abundance of pro-fibrotic proteins.

Chapter 4

Discussion

In this study, the role of miRNAs in regulation of HCC and associated microenvironment development was investigated in the *SRF-VP16*^{iHep} mouse model.

The *SRF-VP16*^{iHep} mice were used to precisely characterize changes in miRNA profile in different stages of hepatocellular carcinoma, from pre-malignant nodular stages to the fully developed HCCs.

Changes in miRNA profile during tumor progression were characterized using sRNA-seq analysis. In parallel, RNA-seq analysis was performed on largely overlapping samples with the aim of characterising mRNA expression profile of *SRF-VP16*^{iHep} HCC, thereby increasing the accuracy of miRNA target predictions.

PCA of obtained sRNA-seq data reveals that with the tumor progression, miRNA dysregulation intensifies and SRF-VP16-driven tumors show distinctly different expression profile compared to nodules and controls. In contrast, PCA of RNA-seq data shows that nodular samples cluster with tumor samples, indicating partially tumor-specific difference in gene expression profiles. This observation is in agreement with the study conducted by Lu et al.[107]. In their study, Lu et al. concluded that miRNA expression allows distinguishing between tumors of different origin as well as between tumors and normal tissue. Furthermore, the study concluded that, unlike with the mRNA expression, a modest number of miRNAs (~200 in total) is sufficient to classify human tumors. Other studies agree with assessment that miRNA profiling can be more accurate at tumor classification since miRNA dysregulation correlates closely with tumor origin and stage[33,108]. Although, this study is performed on relatively small sample size, the sRNA analysis of SRF-VP16-driven tumors supports these conclusions.

GSEA analysis of target genes of upregulated miRNAs, identified using RNA-seq and proteomics, showed strong over-representation of metabolic genes. Pathways identified to be regulated by upregulated miRNA include amino acids-, lipid-, fatty acids-, pyruvate- and Citric Acid (TCA) cycle-related pathways.

Cancer cells in contrast to normal differentiated cells have altered metabolic requirement. As highly proliferative cells, cancer cells adapt their metabolism to facilitate the uptake and incorporation of the nutrients into the biomass as well as production of ATP necessary to generate a new cell[109]. Notably, glucose and pyruvate metabolism are distinctly different in cancers compared to normal tissue. Most differentiated cells metabolize glucose to carbon dioxide by oxidation of pyruvate produced in the glycolysis in the mitohondrial TCA cycle. This reaction produces nicitinamide adenine dinucleotide (NAD^+), reduced (NADH) which is then used in oxidative phosphorylation to maximize adenosine 5'-triphosphate (ATP) production[110]. In contrast, cancer cells convert pyruvate to lactate in anaerobic glycolysis, allowing glycolysis to continue by cycling NADH to NAD^+. Although the process of anaerobic glycolysis is inefficient in terms of ATP production, it provides needed carbon atoms and NADPH necessary for synthesis of amino acids, nucleotides and lipids. This means that glucose together with glutamine supply majority of carbon, nitrogen, free energy and reducing equivalents necessary to support cell growth and division[111]. Glucose is diverted to macromolecular precursors such as acetyl-CoA for fatty acids synthesis, glycolytic intermediates for nonessential amino acids and ribose for nucleotides, while glutamine is channeled in nonessential amino acids production through transamination reactions[111,112].

Majority of pathways identified in GSEA analysis of target genes of upregulated miRNAs are related to metabolism (24 out of 30 statistically significant pathways). Therefore, it is indicative that miRNAs could play important role in metabolic gene downregulation and contribute to the extensive metabolic reprogramming of cancer cells.

GSEA analysis of target genes of downregulated miRNAs profiled using RNA-seq showed strong over-representation of proteins involved in ECM function, integrin signaling and Rho GTPase-related pathways indicating the importance of miRNAs in regulation of the fibrotic microenvironment in HCC development.

The tumor microenvironment is a complex composite of tumor and non-tumor cells embedded within an extracellular matrix (ECM), which facilitates malignant tumor progression[10]. HCC progression is a multistage process that typically arises in the context of liver fibrosis. Fibrosis is the consequence of an exaggerated wound heal-

ing response to chronic liver injury and is characterized by excessive accumulation of ECM[13]. The central event in fibrosis is the activation of hepatic stellate cells (HSCs). Activated HSCs (aHSCs) produce components of the ECM and growth factors, thus causing excessive ECM deposition, neoangiogenesis and inflammation[13]. These processes ultimately result in scarring and thickening of affected tissue, which interferes with normal liver function and facilitates HCC tumorigenesis.

Such a fibrotic microenvironment, caused by quantitative and qualitative changes in ECM depositions, is characterized by increased stiffness, which promotes tumorigenesis through elevated integrin signaling. This signaling leads to enhanced growth, survival and proliferation of tumor cells[20]. Furthermore, ECM deposition enhances HCC chemotherapy resistance and offers protection against immune cells[12].

This study identified a network of functionally connected miRNA:mRNA pairs, which regulate cancer-associated fibrosis. This network regulates different structural, signaling and remodeling ECM components, as well as ECM-linked integrin and Rho-GTPase signaling. Additionally, key miRNA hubs of the network, i.e. miR-29c, miR-335, miR-338, let-7a, let-7c, let-7g, miR-30e and miR-30d are downregulated in the fibrosis-associated mHCC model, two murine fibrosis models and four types of human carcinomas.

miR-29c is a known modulator of fibrotic environments, shown to reduce mRNA levels of different collagens and LAMC1 in nasopharyngeal carcinoma[113]. The role of miR-29c as regulator of fibrosis was re-examined and *Pdgfa, Tpm1, Adatms14* and *Adamts15* were experimentally validated as additional fibrosis-linked targets of this miRNA.

McDaniel et al. found the let-7/Lin28 axis to be involved in human HSC activation upon alcoholic liver injury[114]. Given that Lin28 is neither expressed in SRF-VP16-driven mHCC (Table S1, RNA-seq normalised reads) and pHSC culture model, nor in the RNA-seq dataset of human HSCs of Zhou et al.[115], Lin28-mediated let-7 regulation in pHSCs and mHCC is unlikely. Additionally to Col1a1[114], four different collagen family members, Loxl4 and Tgfbr1 were experimentally validated as let-7 targets.

To my knowledge, none of the other AF-miRNAs, namely miR-335, miR-338, miR-30e and miR-30d, have previously been identified as regulators of fibrosis and none of the AF-miRNAs were previously identified as regulators of fibrotic microenvironments during HCC development.

This study revealed miRNA-mediated targeting of relevant structural components of the ECM, namely laminin and collagens. Collagens represent the most abundant ECM

component and collagen I deposition has been associated with increased incidence of tumor formation and metastasis[105]. A study by Ramaswamy et al.[116] found a gene expression signature that distinguishes primary and metastatic adenocarcinomas and predicts the metastatic probability of tumors. A considerable proportion of the gene-expression signature described by Ramaswamy et al. is comprised of components of the tumor microenvironment, such as *COL1A1* and *COL1A2*.

Additionally to *Col1a1* and *Col1a2*, AF-miRNAs also regulate other collagen subunits, such as *Col5a2*. Although collagen V is a minor constituent of the ECM compared to collagen I, collagen V is essential for fibrillogenesis, as its deletion leads to inability of collagen fibril assembly[117].

Furthermore, AF-miRNAs target components of the PDGF and TGF-β signaling pathways. TGF-β is considered as most potent fibrogenic cytokine. TGF-β binds to type I receptor (Tgfbr1) and causes phosphorylation of downstream SMADs[21], thus inducing SMAD-mediated increased collagen I and III transcription[22]. PDGF is a critical mitogen in the liver, which induces HSC proliferation[17].

In addition to structural and signalling component, AF-miRNAs regulate remodeling components of the ECM. Tropomyosin 1 (Tpm1), an actin-binding protein, helps orienting depositions of collagen and laminin in the ECM[118]. Adamts proteases process procollagens and regulate collagen fibril deposition[119]. *Adamts14* and *Adamts15* are upregulated in SRF-VP16-driven tumors and the CCl_4 mouse model, but downregulated in the pHSC model, possibly indicating the necessity of higher Adamts abundance in tissues, where more extensive ECM remodeling may be required. Loxl proteins are primarily responsible for the regulation of collagen cross-linking. Loxl expression strongly correlates with tumor progression, metastasis and consequently decreased patient survival[120].

Increased ECM stiffness, a consequence of elevated ECM deposition and remodeling, causes activation of MAPKs and Rho-GTPases via integrin signaling. Although experimentally, this study focuses on miRNA-mediated ECM targeting, it is worth mentioning that the identified AF-miRNAs were also predicted to target integrin and Rho-GTPase signaling. These pathways are strong stimulators of tumor migration, invasion and proliferation[121]. $\alpha 5\beta 1$ integrin promotes cell invasion by sensitizing cancer cells to the changes in the ECM[122], while Rho-GTPases are indispensable in the regulation of cell migration and control of multiple aspects of M phase and G_1 progression of the cell cycle[121].

Mechanistically, AF-miRNAs are shown here to be regulated by the transcription factors Pparγ and Egr1. *Pparγ* expression, along with AF-miRNAs, is reduced *in vivo* and *in vitro* upon HSC activation. *Egr1*, however, is significantly downregulated upon pHSC activation in pHSC *in vitro* model, but upregulated in the murine CCl_4 fibrosis model. Overexpression of *Pparg* and *Egr1* lead to increased AF-miRNA expression. Furthermore, Pparγ and Egr1 directly bind to the AF-miRNA-encoding promoters as demonstrated in this study.

PPARs are ligand-inducible transcriptional factors that are members of nuclear hormone receptor superfamily[123]. PPARs are important sensors of cellular levels of fatty acids and fatty-acid derivates. Polyunsaturated fatty acids activate the PPARs with low affinity, while fatty-acid derivates bind more selectivity[124]. Through the mediation of the effects of fatty acids and their derivates at the transcriptional level, PPARs can regulate cell proliferation, differentiation and survival[123]. In mammals there are three PPARs: PPARα, PPARβ/δ and PPARγ. PPARγ exists as two isoforms, PPARγ1 and PPARγ2. While PPARγ2 is restricted to adipose tissue, PPARγ1 is ubiquitously expressed[75].

Pparγ heterodimerizes with the retinoid acid receptor (Rxr) and, upon DNA binding, recruits RNA polymerase and co-activators with histone acetyl transferase activity, causing remodelling of chromatin and enhancing transcription[125]. Pparγ:Rxr interacts with the peroxisome-proliferator response element (PPRE)- usually 5'-AACTAGGNCAAAGGTCA-3' in the promoter of its target genes. PPRE are typically found in various genes involved in lipid metabolism and energy homeostasis[124].

Additionally to its role in regulation of lipid metabolism, Pparγ's involvement in HSC activation has also been described. Pparγ sustains HSC quiescence and promotes de-activation of HSCs[126]. A study by Marra et al.[127] showed that PPARγ inhibits HSC proliferation, migration and chemokine expression, thus inhibiting fibrogenesis. Furthermore, HSC-specific interference of PPARγ signaling aggravates liver damage and fibrosis induced by CCl_4 treatment[15]. Studies conducted so far concluded that PPARγ exerts its anti-fibrotic effects primarily by antagonising TGF-β signaling[119]. Here, I propose an alternative mechanism of Pparγ-mediated regulation of fibrosis via stimulation of anti-fibrotic miRNA expression.

Egr1 recognises highly conserved G-C-rich DNA consensus sequences and activates or represses genes in a zinc-dependent manner[128]. Egr1 is associated with cell growth and differentiation due to its rapid induction upon stimulus[129].

Additionally, as it acts in physiological response to various kinds of stress it is induced during liver injury and controls genes involved in metabolism, cell proliferation and inflammation[128].

However, its role in fibrosis is somewhat controversial. Egr1 has been reported to act either pro-[128] or anti-fibrotically[130,131]. The experiments involving pHSCs and Egr1-overexpressing cells, conducted in this study, showed Egr1 to be anti-fibrotic, while the measurement of *Egr1* expression in CCl_4 model showed Egr1 to be upregulated, therefore acting pro-fibrotic.

Egr1 is required for hepatocyte proliferation after CCl_4 treatment[132]. Since in the CCl_4 experiments conducted in this study whole liver extracts were analysed, it cannot be excluded that increased Egr1 expression is originating from proliferating hepatocytes rather than HSCs.

Alternatively, it is possible that Egr1 is differentially regulated upon different contextual cues, resulting in different functional activities of Egr1.

Activation of HSCs and downregulation of miRNA expression occur in quite a short time (up to seven days) in experimental setup used in this study. Initial miRNA downregulation correlates well with downregulation of the positive regulators of AF-miRNA expression *Pparg* and *Egr1*. However, hypermethylation of let-7a, let-7c-1, miR-335, miR-338 and miR-30e gene promoters upon pHSC activation was also identified. In the window of measurement, the identified methylation changes in all miRNA gene promoters are quite subtle. Therefore, it is possible that initial changes in miRNA expression are mediated through transcription factor regulation, while more long term changes are established through changes in the genomic DNA methylation of miRNA promoters.

The fact that a substantial part of the proteins in the described network is regulated through the activity of such a limited number of miRNAs emphasises the relevance of miRNAs as powerful mediators of complex biological processes, in this case fibrosis.

Chapter 5

Conclusion and Outlook

In the last decade, several individual miRNA:target interactions have been characterized in fibrosis and HCC. However, such approaches ignore the complexity of miRNA signalling networks. Therefore, in this study transcriptome-wide experimental and bioinformatical tools were used to identify miRNA hubs that influence HCC and its fibrotic microenvironment. This approach profiled a complex network of 8 miRNA hubs that target 54 ECM-related genes together regulate structural, signalling and remodelling components of the fibrotic microenvironment.

The findings of this study indicate that the let-7 and miR-30 miRNA families, as well as miR-29c, miR-335 and miR-338 are important anti-fibrotic microRNAs, as these miRNAs are downregulated in the fibrosis-associated mHCC model, two murine fibrosis models and four types of human carcinomas.

Using a multivariate model, which assesses whether examined genes are indeed regulated by miRNAs or by miRNA-unrelated gene regulation mechanisms, this study showed that AF-miRNAs considerably contribute to fibrosis-associated gene regulation *in vivo* in different human fibrosis-facilitated carcinomas.

A noteworthy aspect of the study is the identification of Pparγ and Egr1 as transcription factors which together regulate all miRNA hubs and therefore identify of Pparγ and Egr1 as regulators of a functionally connected class of miRNAs with anti-fibrotic properties.

As all miRNAs are regulated by Pparγ and Egr1, these transcription factors pose attractive targets for anti-fibrotic therapy. Stimulation of Pparγ and/or Egr1 activity using agonist compounds could represent a promising therapy for prevention of fibrosis and HCC in the high risk groups of patients. However, as this study demonstrates that

during activation of HSCs *Pparγ* and *Egr1* are transcriptionally regulated, it is possible that stimulation of activity of these factors may not be sufficient. Therefore, delivery of *Pparγ* and *Egr1* mRNA using lipid nanoparticles might provide a viable alternative.

Additionally, this study indicated important roles of miRNAs in regulation of diverse metabolic pathways. Similar experimental approach as employed to investigate the role of miRNAs in shaping of tumor microenvironment could be used to profile miRNAs hubs that modulate metabolic pathways in HCC. Thus, defining miRNAs contribution to the metabolic reprogramming in HCC development.

Bibliography

[1] Llovet, J. M. *et al.* Hepatocellular carcinoma. *Nat Rev Dis Primers* **2**, 16018 (2016). 1, 3, 5, 6, 7

[2] El-Serag, H. B. Hepatocellular Carcinoma. *New England Journal of Medicine* **365**, 1118–1127 (2011). 1, 2, 3, 4

[3] Forner, A., Reig, M. & Bruix, J. Hepatocellular carcinoma. *Lancet* **391**, 1301–1314 (2018). 1, 2, 3, 4, 6

[4] Yang, J. D. & Roberts, L. R. Hepatocellular carcinoma: a global view. *Nature reviews Gastroenterology & hepatology* **7**, 448 (2010). 1, 2, 3, 4

[5] Balogh, J. *et al.* Hepatocellular carcinoma: a review. *J Hepatocell Carcinoma* **3**, 41–53 (2016). 1, 4

[6] Ally, A. *et al.* Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* **169**, 1327–1341 (2017). 5

[7] Nault, J. C. *et al.* High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. *Nat Commun* **4**, 2218 (2013). 5

[8] Boyault, S. *et al.* Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology* **45**, 42–52 (2007). 5

[9] Totoki, Y. *et al.* Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nat. Genet.* **46**, 1267–1273 (2014). 5

[10] Hernandez-Gea, V., Toffanin, S., Friedman, S. L. & Llovet, J. M. Role of the microenvironment in the pathogenesis and treatment of hepatocellular carcinoma. *Gastroenterology* **144**, 512–527 (2013). 6, 7, 8, 9, 10, 92, 108

[11] Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011). 6

[12] Heindryckx, F. & Gerwins, P. Targeting the tumor stroma in hepatocellular carcinoma. *World J Hepatol* **7**, 165–176 (2015). 7, 10, 109

[13] Hernandez-Gea, V. & Friedman, S. L. Pathogenesis of liver fibrosis. *Annu Rev Pathol* **6**, 425–456 (2011). 9, 10, 11, 12, 109

[14] Friedman, S. L. Hepatic stellate cells: protean, multifunctional, and enigmatic cells of the liver. *Physiol. Rev.* **88**, 125–172 (2008). 9, 44, 64, 74, 77, 82

[15] Tsuchida, T. & Friedman, S. L. Mechanisms of hepatic stellate cell activation. *Nat Rev Gastroenterol Hepatol* **14**, 397–411 (2017). 9, 11, 111

[16] Bataller, R. & Brenner, D. A. Liver fibrosis. *J. Clin. Invest.* **115**, 209–218 (2005). 9

[17] Pellicoro, A., Ramachandran, P., Iredale, J. P. & Fallowfield, J. A. Liver fibrosis and repair: immune regulation of wound healing in a solid organ. *Nat. Rev. Immunol.* **14**, 181–194 (2014). 10, 11, 12, 110

[18] Carloni, V., Luong, T. V. & Rombouts, K. Hepatic stellate cells and extracellular matrix in hepatocellular carcinoma: more complicated than ever. *Liver Int.* **34**, 834–843 (2014). 10

[19] Schrader, J. *et al.* Matrix stiffness modulates proliferation, chemotherapeutic response, and dormancy in hepatocellular carcinoma cells. *Hepatology* **53**, 1192–1205 (2011). 10

[20] Kalluri, R. & Zeisberg, M. Fibroblasts in cancer. *Nat. Rev. Cancer* **6**, 392–401 (2006). 10, 77, 109

[21] Derynck, R. & Zhang, Y. E. Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature* **425**, 577–584 (2003). 10, 110

[22] Mann, D. A. & Marra, F. Fibrogenic signalling in hepatic stellate cells. *J. Hepatol.* **52**, 949–950 (2010). 10, 110

[23] Lee, Y. A., Wallace, M. C. & Friedman, S. L. Pathobiology of liver fibrosis: a translational success story. *Gut* **64**, 830–841 (2015). 11, 17, 23

[24] Jun, J. I. & Lau, L. F. Resolution of organ fibrosis. *J. Clin. Invest.* **128**, 97–107 (2018). 11, 12

[25] Kisseleva, T. *et al.* Myofibroblasts revert to an inactive phenotype during regression of liver fibrosis. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 9448–9453 (2012). 12

[26] Ghildiyal, M. & Zamore, P. D. Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* **10**, 94–108 (2009). 12

[27] Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009). 12, 13

[28] Pasquinelli, A. E. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat. Rev. Genet.* **13**, 271–282 (2012). 12, 13

[29] Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19**, 92–105 (2009). 13

[30] Eulalio, A. *et al.* Deadenylation is a widespread effect of miRNA regulation. *RNA* **15**, 21–32 (2009). 13

[31] Hausser, J. & Zavolan, M. Identification and consequences of miRNA-target interactions–beyond repression of gene expression. *Nat. Rev. Genet.* **15**, 599–612 (2014). 13

[32] Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 509–524 (2014). 13, 14, 15

[33] Lujambio, A. & Lowe, S. W. The microcosmos of cancer. *Nature* **482**, 347–355 (2012). 14, 17, 19, 107

[34] Bracken, C. P., Scott, H. S. & Goodall, G. J. A network-biology perspective of microRNA function and dysfunction in cancer. *Nat. Rev. Genet.* **17**, 719–732 (2016). 14

[35] Winter, J., Jung, S., Keller, S., Gregory, R. I. & Diederichs, S. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat. Cell Biol.* **11**, 228–234 (2009). 14, 16

[36] Yates, L. A., Norbury, C. J. & Gilbert, R. J. The long and short of microRNA. *Cell* **153**, 516–519 (2013). 14

[37] Ramalingam, P. *et al.* Biogenesis of intronic miRNAs located in clusters by independent transcription and alternative splicing. *RNA* **20**, 76–87 (2014). 14

[38] Krol, J., Loedige, I. & Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.* **11**, 597–610 (2010). 15, 17

[39] Suzuki, H., Maruyama, R., Yamamoto, E. & Kai, M. Epigenetic alteration and microRNA dysregulation in cancer. *Front Genet* **4**, 258 (2013). 15

[40] Lee, E. J. *et al.* Systematic evaluation of microRNA processing patterns in tissues, cell lines, and tumors. *RNA* **14**, 35–42 (2008). 15

[41] Davis, B. N., Hilyard, A. C., Lagna, G. & Hata, A. SMAD proteins control DROSHA-mediated microRNA maturation. *Nature* **454**, 56–61 (2008). 15

[42] Suzuki, H. I. *et al.* Modulation of microRNA processing by p53. *Nature* **460**, 529–533 (2009). 15

[43] Michlewski, G. & Caceres, J. F. Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. *Nat. Struct. Mol. Biol.* **17**, 1011–1018 (2010). 15

[44] Trabucchi, M. *et al.* The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* **459**, 1010–1014 (2009). 15

[45] Kim, Y. K., Heo, I. & Kim, V. N. Modifications of small RNAs and their associated proteins. *Cell* **143**, 703–709 (2010). 15

[46] Ameres, S. L. & Zamore, P. D. Diversifying microRNA sequence and function. *Nat. Rev. Mol. Cell Biol.* **14**, 475–488 (2013). 17

[47] Kawahara, Y., Zinshteyn, B., Chendrimada, T. P., Shiekhattar, R. & Nishikura, K. RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. *EMBO Rep.* **8**, 763–769 (2007). 17

[48] Cloonan, N. *et al.* MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol.* **12**, R126 (2011). 17

[49] Ameres, S. L. *et al.* Target RNA-directed trimming and tailing of small silencing RNAs. *Science* **328**, 1534–1539 (2010). 17

[50] Heo, I. *et al.* Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol. Cell* **32**, 276–284 (2008). 17

[51] Heo, I. *et al.* Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. *Cell* **151**, 521–532 (2012).

[52] Nam, Y., Chen, C., Gregory, R. I., Chou, J. J. & Sliz, P. Molecular basis for interaction of let-7 microRNAs with Lin28. *Cell* **147**, 1080–1091 (2011). 17

[53] Ruegger, S. & Grosshans, H. MicroRNA turnover: when, how, and why. *Trends Biochem. Sci.* **37**, 436–446 (2012). 17

[54] Callegari, E. *et al.* MicroRNAs in liver cancer: a model for investigating pathogenesis and novel therapeutic approaches. *Cell Death Differ.* **22**, 46–57 (2015). 17

[55] Thomson, J. M. *et al.* Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev.* **20**, 2202–2207 (2006). 17

[56] Kumar, M. S., Lu, J., Mercer, K. L., Golub, T. R. & Jacks, T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat. Genet.* **39**, 673–677 (2007). 17

[57] Palanichamy, J. K. & Rao, D. S. miRNA dysregulation in cancer: towards a mechanistic understanding. *Front Genet* **5**, 54 (2014). 17

[58] Olson, E. N. & Nordheim, A. Linking actin dynamics and gene transcription to drive cellular motile functions. *Nat. Rev. Mol. Cell Biol.* **11**, 353–365 (2010). 18

[59] Posern, G. & Treisman, R. Actin' together: serum response factor, its cofactors and the link to signal transduction. *Trends Cell Biol.* **16**, 588–596 (2006). 18

[60] Clark, K. A. & Graves, B. J. Dual views of SRF: a genomic exposure. *Genes Dev.* **28**, 926–928 (2014). 18

[61] Wei, G. H. *et al.* Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *EMBO J.* **29**, 2147–2160 (2010). 18, 128

[62] Esnault, C. *et al.* Rho-actin signaling to the MRTF coactivators dominates the immediate transcriptional response to serum in fibroblasts. *Genes Dev.* **28**, 943–958 (2014). 18

[63] Medjkane, S., Perez-Sanchez, C., Gaggioli, C., Sahai, E. & Treisman, R. Myocardin-related transcription factors and SRF are required for cytoskeletal dynamics and experimental metastasis. *Nat. Cell Biol.* **11**, 257–268 (2009). 18

[64] Ohrnberger, S. *et al.* Dysregulated serum response factor triggers formation of hepatocellular carcinoma. *Hepatology* **61**, 979–989 (2015). 18, 21, 22

[65] Anastasiadou, E., Jacob, L. S. & Slack, F. J. Non-coding RNA networks in cancer. *Nat. Rev. Cancer* **18**, 5–18 (2018). 19

[66] Dalton, S. & Treisman, R. Characterization of SAP-1, a protein recruited by serum response factor to the c-fos serum response element. *Cell* **68**, 597–612 (1992). 21

[67] Sandstrom, J. *et al.* Degeneration of the mouse retina upon dysregulated activity of serum response factor. *Mol. Vis.* **17**, 1110–1127 (2011). 21

[68] Scholten, D., Trebicka, J., Liedtke, C. & Weiskirchen, R. The carbon tetrachloride model in mice. *Lab. Anim.* **49**, 4–11 (2015). 22, 82

[69] Borojevic, R. *et al.* Establishment of a continuous cell line from fibrotic schistosomal granulomas in mice livers. *In Vitro Cell. Dev. Biol.* **21**, 382–390 (1985). 22

[70] Margis, R. & Borojevic, R. Retinoid-mediated induction of the fat-storing phenotype in a liver connective tissue cell line (GRX). *Biochim. Biophys. Acta* **1011**, 1–5 (1989). 22

[71] TODARO, G. J. & GREEN, H. Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. *J. Cell Biol.* **17**, 299–313 (1963). 23

[72] Negmadjanov, U. *et al.* TGF-β-mediated differentiation of fibroblasts is associated with increased mitochondrial content and cellular respiration. *PLoS ONE* **10**, e0123046 (2015). 23

[73] Weiskirchen, S., Tag, C. G., Sauer-Lehnen, S., Tacke, F. & Weiskirchen, R. Isolation and Culture of Primary Murine Hepatic Stellate Cells. *Methods Mol. Biol.* **1627**, 165–191 (2017). 23, 82

[74] Viswanathan, S. R. *et al.* Lin28 promotes transformation and is associated with advanced human malignancies. *Nat. Genet.* **41**, 843–848 (2009). 24

[75] Tontonoz, P., Hu, E., Graves, R. A., Budavari, A. I. & Spiegelman, B. M. mPPAR gamma 2: tissue-specific regulator of an adipocyte enhancer. *Genes Dev.* **8**, 1224–1234 (1994). 24, 111

[76] Yu, J., de Belle, I., Liang, H. & Adamson, E. D. Coactivating factors p300 and CBP are transcriptionally crossregulated by Egr1 in prostate cells, leading to divergent responses. *Mol. Cell* **15**, 83–94 (2004). 24

[77] Potter, H. & Heller, R. Transfection by Electroporation. *Curr Protoc Mol Biol* **121**, 1–9 (2018). 42

[78] Sambrook, J. & Russell, D. W. DNA transfection by electroporation. *CSH Protoc* **2006** (2006). 43

[79] Felgner, P. L. *et al.* Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7413–7417 (1987). 43

[80] Wang, Z. The guideline of the design and validation of MiRNA mimics. *Methods Mol. Biol.* **676**, 211–223 (2011). 43

[81] Glasel, J. A. Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. *BioTechniques* **18**, 62–63 (1995). 49

[82] Krebs, S., Fischaleck, M. & Blum, H. A simple and loss-free method to remove TRIzol contaminations from minute RNA samples. *Anal. Biochem.* **387**, 136–138 (2009). 49

[83] Jones, L. J., Yue, S. T., Cheung, C. Y. & Singer, V. L. RNA quantitation by fluorescence-based solution assay: RiboGreen reagent characterization. *Anal. Biochem.* **265**, 368–374 (1998). 49

[84] Ruijter, J. M. *et al.* Evaluation of qPCR curve analysis methods for reliable biomarker discovery: bias, resolution, precision, and implications. *Methods* **59**, 32–46 (2013). 52, 53

[85] Ramakers, C., Ruijter, J. M., Deprez, R. H. & Moorman, A. F. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci. Lett.* **339**, 62–66 (2003). 53

[86] Karlen, Y., McNair, A., Perseguers, S., Mazza, C. & Mermod, N. Statistical significance of quantitative PCR. *BMC Bioinformatics* **8**, 131 (2007). 53

[87] Daniel, B., Balint, B. L., Nagy, Z. S. & Nagy, L. Mapping the genomic binding sites of the activated retinoid X receptor in murine bone marrow-derived macrophages using chromatin immunoprecipitation sequencing. *Methods Mol. Biol.* **1204**, 15–24 (2014). 54

[88] Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). 56, 58, 71, 131

[89] Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). 58, 72, 131

[90] Marcel, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* **17**, 10–12 (2011). 58, 131

[91] Ehrich, M. *et al.* Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15785–15790 (2005). 59

[92] Reczko, M., Maragkakis, M., Alexiou, P., Grosse, I. & Hatzigeorgiou, A. G. Functional microRNA targets in protein coding sequences. *Bioinformatics* **28**, 771–776 (2012). 64, 74

[93] Paraskevopoulou, M. D. *et al.* DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Res.* **41**, W169–173 (2013). 64, 74

[94] Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4** (2015). 64, 74

[95] Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000). 64, 74

[96] Fabregat, A. *et al.* Reactome graph database: Efficient access to complex pathway data. *PLoS Comput. Biol.* **14**, e1005968 (2018). 64, 74

[97] Liu, Q. *et al.* Identification of active miRNA promoters from nuclear run-on RNA sequencing. *Nucleic Acids Res.* **45**, e121 (2017). 66, 99

[98] Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008). 66, 99

[99] Engstrom, P. G. *et al.* Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat. Methods* **10**, 1185–1191 (2013). 67

[100] Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010). 67

[101] Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011). 67, 99

[102] Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W. W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* **32**, D91–94 (2004). 67

[103] Jacobsen, A. *et al.* Analysis of microRNA-target interactions across diverse cancer types. *Nat. Struct. Mol. Biol.* **20**, 1325–1332 (2013). 68, 92

[104] Miralles, F., Posern, G., Zaromytidou, A. I. & Treisman, R. Actin dynamics control SRF activity by regulation of its coactivator MAL. *Cell* **113**, 329–342 (2003). 77

[105] Gilkes, D. M., Semenza, G. L. & Wirtz, D. Hypoxia and the extracellular matrix: drivers of tumour metastasis. *Nat. Rev. Cancer* **14**, 430–439 (2014). 79, 110

[106] Newman, M. A., Thomson, J. M. & Hammond, S. M. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA* **14**, 1539–1549 (2008). 86

[107] Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435**, 834–838 (2005). 107

[108] van Kouwenhove, M., Kedde, M. & Agami, R. MicroRNA regulation by RNA-binding proteins and its implications for cancer. *Nat. Rev. Cancer* **11**, 644–656 (2011). 107

[109] Sullivan, L. B., Gui, D. Y. & Vander Heiden, M. G. Altered metabolite levels in cancer: implications for tumour biology and cancer therapy. *Nat. Rev. Cancer* **16**, 680–693 (2016). 108

[110] Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029–1033 (2009). 108

[111] Altman, B. J., Stine, Z. E. & Dang, C. V. From Krebs to clinic: glutamine metabolism to cancer therapy. *Nat. Rev. Cancer* **16**, 619–634 (2016). 108

[112] Rohrig, F. & Schulze, A. The multifaceted roles of fatty acid synthesis in cancer. *Nat. Rev. Cancer* **16**, 732–749 (2016). 108

[113] Sengupta, S. *et al.* MicroRNA 29c is down-regulated in nasopharyngeal carcinomas, up-regulating mRNAs encoding extracellular matrix proteins. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5874–5878 (2008). 109

[114] McDaniel, K. *et al.* The let-7/Lin28 axis regulates activation of hepatic stellate cells in alcoholic liver injury. *J. Biol. Chem.* **292**, 11336–11347 (2017). 109

[115] Zhou, C. *et al.* Long noncoding RNAs expressed in human hepatic stellate cells form networks with extracellular matrix proteins. *Genome Med* **8**, 31 (2016). 109

[116] Ramaswamy, S., Ross, K. N., Lander, E. S. & Golub, T. R. A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* **33**, 49–54 (2003). 110

[117] Mouw, J. K., Ou, G. & Weaver, V. M. Extracellular matrix assembly: a multiscale deconstruction. *Nat. Rev. Mol. Cell Biol.* **15**, 771–785 (2014). 110

[118] Hayes, A. J., Benjamin, M. & Ralphs, J. R. Role of actin stress fibres in the development of the intervertebral disc: cytoskeletal control of extracellular matrix assembly. *Dev. Dyn.* **215**, 179–189 (1999). 110

[119] Bonnans, C., Chou, J. & Werb, Z. Remodelling the extracellular matrix in development and disease. *Nat. Rev. Mol. Cell Biol.* **15**, 786–801 (2014). 110, 111

[120] Barker, H. E., Cox, T. R. & Erler, J. T. The rationale for targeting the LOX family in cancer. *Nat. Rev. Cancer* **12**, 540–552 (2012). 110

[121] Jaffe, A. B. & Hall, A. Rho GTPases: biochemistry and biology. *Annu. Rev. Cell Dev. Biol.* **21**, 247–269 (2005). 110

[122] Hamidi, H. & Ivaska, J. Every step of the way: integrins in cancer progression and metastasis. *Nat. Rev. Cancer* **18**, 533–548 (2018). 110

[123] Ahmadian, M. *et al.* PPARg signaling and metabolism: the good, the bad and the future. *Nat. Med.* **19**, 557–566 (2013). 111

[124] Michalik, L., Desvergne, B. & Wahli, W. Peroxisome-proliferator-activated receptors and cancers: complex stories. *Nat. Rev. Cancer* **4**, 61–70 (2004). 111

[125] Peters, J. M., Shah, Y. M. & Gonzalez, F. J. The role of peroxisome proliferator-activated receptors in carcinogenesis and chemoprevention. *Nat. Rev. Cancer* **12**, 181–195 (2012). 111

[126] Michalik, L. & Wahli, W. Involvement of PPAR nuclear receptors in tissue injury and wound repair. *J. Clin. Invest.* **116**, 598–606 (2006). 111

[127] Marra, F. *et al.* Ligands of peroxisome proliferator-activated receptor gamma modulate profibrogenic and proinflammatory actions in hepatic stellate cells. *Gastroenterology* **119**, 466–478 (2000). 111

[128] Magee, N. & Zhang, Y. Role of early growth response 1 in liver metabolism and liver cancer. *Hepatoma Res* **3**, 268–277 (2017). 111, 112

[129] Yan, S. F., Pinsky, D. J., Mackman, N. & Stern, D. M. Egr-1: is it always immediate and early? *J. Clin. Invest.* **105**, 553–554 (2000). 111

[130] Khimji, A. K., Shao, R. & Rockey, D. C. Divergent transforming growth factor-beta signaling in hepatic stellate cells after liver injury: functional effects on ECE-1 regulation. *Am. J. Pathol.* **173**, 716–727 (2008). 112

[131] Sullivan, B. P., Cui, W., Copple, B. L. & Luyendyk, J. P. Early growth response factor-1 limits biliary fibrosis in a model of xenobiotic-induced cholestasis in mice. *Toxicol. Sci.* **126**, 267–274 (2012). 112

[132] Pritchard, M. T., Malinak, R. N. & Nagy, L. E. Early growth response (EGR)-1 is required for timely cell-cycle entry and progression in hepatocytes after acute carbon tetrachloride exposure in mice. *Am. J. Physiol. Gastrointest. Liver Physiol.* **300**, G1124–1131 (2011). 112

[133] Low, K. C. & Tergaonkar, V. Telomerase: central regulator of all of the hallmarks of cancer. *Trends Biochem. Sci.* **38**, 426–434 (2013). 127

[134] Huang, F. W. *et al.* Highly recurrent TERT promoter mutations in human melanoma. *Science* **339**, 957–959 (2013). 127

[135] Borah, S. *et al.* Cancer. TERT promoter mutations and telomerase reactivation in urothelial cancer. *Science* **347**, 1006–1010 (2015). 127

[136] Horn, S. *et al.* TERT promoter mutations in familial and sporadic melanoma. *Science* **339**, 959–961 (2013). 127

[137] Rachakonda, P. S. *et al.* TERT promoter mutations in bladder cancer affect patient survival and disease recurrence through modification by a common polymorphism. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17426–17431 (2013). 127

[138] Hollenhorst, P. C. *et al.* Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. *Genes Dev.* **25**, 2147–2157 (2011). 128

[139] Bell, R. J. *et al.* Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. *Science* **348**, 1036–1039 (2015). 128

[140] Denisova, E. *et al.* Frequent DPH3 promoter mutations in skin cancers. *Oncotarget* **6**, 35922–35930 (2015). 128

[141] Liu, S. *et al.* Dph3, a small protein required for diphthamide biosynthesis, is essential in mouse development. *Mol. Cell. Biol.* **26**, 3835–3841 (2006). 128

[142] Fichtner, L. *et al.* Elongator's toxin-target (TOT) function is nuclear localization sequence dependent and suppressed by post-translational modification. *Mol. Microbiol.* **49**, 1297–1307 (2003). 128

[143] Wang, L. *et al.* Silencing of diphthamide synthesis 3 (Dph3) reduces metastasis of murine melanoma. *PLoS ONE* **7**, e49988 (2012). 128

[144] Macian, F. NFAT proteins: key regulators of T-cell development and function. *Nat. Rev. Immunol.* **5**, 472–484 (2005). 129

[145] Muller, M. R. & Rao, A. NFAT, immunity and cancer: a transcription factor comes of age. *Nat. Rev. Immunol.* **10**, 645–656 (2010). 129

[146] Michel, C. I. *et al.* Small nucleolar RNAs U32a, U33, and U35a are critical mediators of metabolic stress. *Cell Metab.* **14**, 33–44 (2011). 130

[147] Sabarinathan, R., Mularoni, L., Deu-Pons, J., Gonzalez-Perez, A. & Lopez-Bigas, N. Nucleotide excision repair is impaired by binding of transcription factors to DNA. *Nature* **532**, 264–267 (2016). 130

[148] Perera, D. *et al.* Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes. *Nature* **532**, 259–263 (2016). 130

[149] Tomfohr, J., Lu, J. & Kepler, T. B. Pathway level analysis of gene expression using singular value decomposition. *BMC Bioinformatics* **6**, 225 (2005). 132

[150] Yang, M. & Vousden, K. H. Serine and one-carbon metabolism in cancer. *Nat. Rev. Cancer* **16**, 650–662 (2016). 132

[151] Locasale, J. W. Serine, glycine and one-carbon units: cancer metabolism in full circle. *Nat. Rev. Cancer* **13**, 572–583 (2013). 132

[152] Kinnaird, A., Zhao, S., Wellen, K. E. & Michelakis, E. D. Metabolic control of epigenetics in cancer. *Nat. Rev. Cancer* **16**, 694–707 (2016). 132

Appendix A

Summary of other projects

During my PhD work I have been involved in three additional projects. As these projects deal with topics distinct from the main topic (i.e. function of miRNAs in murine liver carcinogenesis) I will provide here a brief summary of the topics and findings of these projects.

A.1 NonComs: Non-coding mutations in cancer

This study investigated the mechanisms of the influence of point mutations in the gene promoters on transcriptional dysregulation. The main aim of the study was to determine if the investigated mutations generated or destroyed a transcription factor binding site and if so, to identify the transcription factor whose binding site was influenced by the mutation.

Initially the study focused on mutations in the *TERT* gene and was later expanded to *DPH3, RPL13A, RPS20, RNF185, GMCL1* and *PES1* genes.

Maintenance of telomere repeats at the end of chromosomes is the prime function of the telomerase holoenzyme, which is encoded by the TERT gene. Interestingly, recent studies have implicated telomerase activity in regulation of the various hallmarks of cancer, including genome instability, transcriptional regulation and metabolic reprogramming[133]. Several studies have identified and confirmed two highly recurrent mutations in the promoter of the human TERT (hTERT) gene in different human carcinomas[134–136]. Mutations were correlated with the increase of hTERT transcription and enzyme activity. Elevated enzyme activity was correlated with reduced disease-specific survival[135,137].

These two independent mutations generate a binding motif for ETS transcription factors, which was postulated to be causative for hTERT upregulation. Members of the large ETS family of transcription factors have diverse functions and activities in physiology and oncogenesis. ETS factors play a role in angiogenesis, hematopoiesis and neuronal development, and activated ETS factors induce cell proliferation, differentiation and migration[138]. The ETS family of transcription factors contains 28 members. Although all ETS factors have highly conserved core DNA binding domains, different ETS proteins may exhibit a preference for different flanking sequences and bind differentially to specific DNA sites[61].

As ETS binding was proposed based on a bioinformatical analysis, the aim was to investigate ETS binding to mutated hTERT promoter regions experimentally. To explore ETS binding to the TERT promoter region the ChIP method was used.

Unfortunately, during the ChIP optimisation procedure a competitor study Bell et al.[139] had identified the GA Binding Protein Transcription Factor, Alpha (GABPA) as an ETS factor that binds to the promoter of the TERT gene[139]. The identification of binding of this ETS factor in the TERT promoter has shifted my attention to the human DPH3 gene[140].

Diphthamide biosynthesis 3 (DPH3) is one of five proteins (DPH1 to 5) involved in diphthamide synthesis, a modified version of histidine which is incorporated at AA position 715 of elongation factor-2 (eEF-2)[141]. Additionally, DPH3 was shown to associate with the elongator complex in yeast, thereby regulating its activity during transcription elongation[142]. Involvement of DPH3 in murine melanoma metastasis *in vivo* and *in vitro* has been shown by Lei Wang et al.[143]. Similar to *hTERT*, the *DPH3* gene promoter contains mutations close to the *DPH3* TSS in human tumors. These mutations were found in skin cancers: melanoma (frequency of 10%), basal cell carcinoma (42%) and squamous cell carcinoma (39%). *In silico* analysis indicated strong binding of ETS transcription factors to the wild type promoter and non-optimal ETS protein binding to the mutated promoter.

To characterize the effect of the mutations on *DPH3* expression, *DPH3* transcript expression was compared in paired DPH3 mutant and wild type squamous cell carcinoma cell lines and paired melanoma cell lines. The expression analysis showed that mutant cell lines expressed a higher level of *DPH3* protein coding transcripts. The same paired cell lines where used to explore ETS binding to the *DPH3* promoter region using ChIP. ChIP analysis showed strong binding of GABPA and E74 Like ETS Transcription Factor 1 (ELF1) to the locus and weak binding of ELK1 and ELK4. However, analysis showed

that none of these factors bind differentially to the mutated promoter in comparison to the wild type promoter. As the *DPH3* mutation in the mutant cell lines is heterozygous, it is possible that the wild-type *DPH3* allele is masking the effect of the mutation on ETS factor binding. To exclude this possibility, imunoprecipitated and input samples from the ChIP analysis were sequenced. Sequencing electropherograms showed a difference in binding of RNA polymerase II (PolII) and GABPA to the mutated promoter. To test if the presence of the mutation influences chromatin modifications, ChIP quantifying H3K4me2 and H3K27me3 signals was performed. There was no difference in the signal quantity of H3K4me2 and H3K27me3 in paired mutant/wild type cell lines as determined by ChIP coupled with qPCR or sequencing analysis of imunoprecipitated samples.

To profile further potential candidates for immunoprecipitation, we employed protein pull-down strategies and subsequent mass-spectrometry protein identification in collaboration with the group of M. Vermeulen (Radboud University, Netherlands). The most promising identified candidates with mutation-specific alterations in binding to the mutated sites were Nuclear factor of activated T cells (NFAT) transcription factors.

NFATs were originally identified as major regulators of T cell development[144]. Since their discovery more then two decades ago, intensive research showed their important function in many developmental programs in vertebrates, including heart, skeletal muscle, blood vessels, neurons, pancreas, bone and skin[145]. Furthermore, numerous studies have implicated hyperactive or overexpressed NFATs in tumor development and metastasis formation[145]. NFATs are regulated by intracellular Ca^{2+} signaling which leads to nuclear translocation[144]. Immunostaining showed that in the cell lines used in expression, ChIP and pull-down analysis, activation of Ca^{2+} signaling using ionophores is necessary to trigger NFAT translocation. In unstimulated cells NFAT1 was completely absent from the nucleus, arguing against the NFAT involvement in *DPH3* overexpression. Furthermore, ChIP analysis failed to detected NFAT1 and NFAT2 binding to the *DPH3* promoter.

A nucleosome scanning assay showed presence of nucleosome at the position of the mutation site in DPH3 mutant and wild-type cell lines, excluding the possibility of nucleosome displacement by the mutated site.

As part of the effort of the NonCoM Consortium, whole gene sequencing was performed on 40 melanoma samples together with corresponding control samples. This analysis, performed by the group of Computational Oncology at DKFZ, profiled 21 mutations in the non coding part of the genome which occurred with a frequency higher

the 10%. We focused on 5 mutation sites in the promoters of the genes *ribosomal protein L13a (RPL13A), ribosomal protein S20 (RPS20), ring finger protein 185 (RNF185), germ cell-less, spermatogenesis associated 1 (GMCL1)* and *pescadillo ribosomal biogenesis factor 1 (PES1)*.

In silico analysis indicated differential binding of ETS transcription factors to the wild-type promoter compared to the mutated promoter in all aforementioned cases.

To characterize the effect of the mutations on gene expression, expression of the previously listed genes was compared in paired mutant and wild-type melanoma cell lines. The expression analysis showed that there was no difference in gene expression between mutant and wild-type cell lines. The same paired cell lines where used to explore ETS binding to the gene promoter regions using ChIP. ChIP analysis showed strong binding of GABPA and weak binding of Elk1 and Elk4 to the promoters of all genes analysed. However, analysis showed that none of these factors bind differentially to the mutated promoter in comparison to the wild type promoter. As the mutations in the mutant cell lines are heterozygous, it is possible that, as in the case of *DPH3*, the wild-type allele is masking the effect of the mutation on ETS factor binding. To exclude this possibility, imunoprecipitated and input samples from the ChIP analysis were sequenced. Sequencing electropherograms showed a difference in binding of GABPA and H3K27me3 to the mutated promoter of *RPL13A*. GABPA preferentially binds to the wild-type allele while, in correspondence, the H3K27me3 signal is more prominent on the mutant allele. There was no difference in the signal quantity of PolII and H3K4me2 in paired mutant/wild type cell lines as determined by ChIP coupled with qPCR or sequencing analysis of imunoprecipitated samples. Furthermore, there is no difference in the signal quantity of GABPA, PolII, H3K4me2 and H3K27me3 in the promoters of the other listed genes between paired mutant/wild type cell lines as determined by sequencing analysis of imunoprecipitated samples.

The *RPL13A* gene contains in its introns 4 small nucleolar RNAs (snoRNA) which, according to the study performed by Michel et al.[146], are critical modulators of oxidative stress. Therefore, it is possible that modulation of the GABPA binding to the promoter of this gene is associated with oxidative stress response in tumors.

Recently, studies by Perera et al. and Sabarinathan et al.[147,148] concluded that the rate of somatic mutations in melanomas is highly increased at active transcription factor binding sites and nucleosome embedded DNA due to decreased nucleotide excision repair activity at these sites.

The results obtained in this study indicate that these mutations do not significantly influence overall binding of ETS factors to the mutated site, possibly due to degenerative binding of ETS factors.

In summary, this study experimentally showed that all of the studied mutations are most likely passenger mutations, i.e. mutations that do not provide a growth advantage to the tumor.

A.2 Characterization of the LT cell lines

To address functional and regulatory aspects of miRNA activity and other biological processes, cell lines, named LT1, LT2 and LT3, were established from three independent SRF-VP16-driven mouse tumors. The primary cell line LT1 was successfully cloned in two distinct subcloned cell lines (LT1.1 and LT1.3), LT2 in three subcloned cell lines (LT2.1, LT2.3 and LT2.5) and LT3 into three subcloned cell lines (LT3.1, LT3.2 and LT3.3). The primary aim of this study was the characterization of different molecular and cellular aspects of these cell lines. To characterize the growth rates of aforementioned cell lines, the growth curves of the cell lines were determined. Proliferation of the cell lines was also tested in different media in order to formulate the optimal growth conditions for the cell lines. Among the LT cell lines, LT2.1 and LT2.5 have the fastest proliferation rates, while LT1.1 has the slowest. All cell lines were compared in clonogenic assays and their migratory potential was assessed. LT2.1 showed the highest migratory potential and the highest colony formation ability.

Karyotyping of the cell line was performed in collaboration with Prof. Krämer's group (DKFZ). LT2 cell lines show greater genome instability, characterized with a large number of events of chromosomes aneuploidies, duplications and deletions, compared to LT1 cell lines.

RNA-seq and sRNA-seq analysis of the cell lines was performed in collaboration with Dr. Geffer's group (Helmholtz center, Braunschweig). RNA-seq and sRNA-seq data of the cell lines were compared to the already generated RNA-seq mHCC data. FASTQC tool (`https://www.bioinformatics.babraham.ac.uk/projects/fastqc/`) was used to analyse the quality of the raw reads and Cutadapt[90] to trim the adapters at the reads ends. To map and count the reads STAR[88] mapper was used. Differential gene expression analysis was performed using DESeq2[89]. RNA-seq and sRNA-seq analysis showed that the LT2.1 cell line has a distinct transcription profile in comparison

to the other cell lines, which showed similar expression profiles. The transcriptome and miRNome data are in agreement with our earlier cellular characterization, which showed significantly faster growth and migratory potential of LT2.1 in comparison to the other cell lines. Single sample gene set enrichment analysis[149] has shown different activation levels of metabolic and signaling pathways in LT2.1 in comparison to the other cell lines. Comparison of cell line RNA-seq data to the tumor RNA-seq data and consequent gene set enrichment analysis has shown that most of the genes that show upregulation or downregulation in both cell lines and tumors are metabolic genes. The regulation of genes that are part of signaling pathways does not show conservation between cell lines and tissue. Analysis of the miRNAs and their targets showed that miRNAs which show similar dysregulation between cell lines and tumors target mainly metabolic genes. Altered metabolic features reflected in transcriptome changes primarily include reactions that support altered and enhanced bioenergetics, biosynthesis and redox balance.

Considering the similarity of metabolic pathway regulation and expression between mHCC and cell lines derived from these tumors, the third project described here focuses on the investigation of one carbon pool by folate and connected serine, glycine metabolism and nucleotide synthesis pathways in mHCC and LT cell lines.

A.3 Metabolic reprogramming of serine synthesis, mitochondrial one-carbon metabolism, and methionine cycle activity in hepatocellular carcinoma

Folate one-carbon metabolism contributes to nucleotide synthesis, methylation reactions and the generation of Nicotinamide adenine dinucleotide phosphate, reduced (NAPDH) for reactive oxygen species defense[150]. One-carbon metabolism is mirrored in the cytoplasm and the mitochondria, where closely related but distinct enzymes use one-carbon units derived from serine and glycine to generate folate intermediates which can be utilised in *de novo* purine synthesis, thymidilate synthesis and the remethylation of homocysteine[151]. Through remethylation of homocysteine, one-carbon metabolism contributes to S-adenosylmethionine (SAM) production. SAM serves as a main donor of methyl groups to variuos biomolecules, including proteins, nucleic acids, lipids and secondary metabolites[152].

In collaboration with the group of Prof. Rabinowitz (Princeton University) a metabolomics analysis of LT cell lines was performed. The generated results support the transcriptomic analysis and show a difference in metabolic flux between LT2.1 and LT2.5. Namely, one-carbon metabolism in LT2.5 cells in mostly directed through the mitochondrial one-carbon branch, while LT2.1 uses both the mitochondrial and cytoplasmic branches. Connected with the changes in one-carbon metabolism, western blots have shown a difference in global histone methylation between LT2.1 and LT2.5 cell lines.

Appendix B

Abbreviations

ACTA2 Alpha Smooth Muscle Actin

ADARs Adenosine deaminases

AF-miRNAs anti-fibrotic miRNAs

APC adenomatous polyposis coli

ATP adenosine 5'-triphosphate

AXIN1 axin 1

Acta2 alpha-smooth muscle actin

Adamts A Disintegrin and Metalloproteinases with a ThromboSpondin motif

Ago Argonaute

BAFs RG1- or HRBM-associated factors

BCL-2 B-cell lymphoma 2

BCLC Barcelona Clinic Liver Cancer

BRCA breast invasive carcinoma

BSA Albumin

CCl_4 carbon tetrachlorid

CIAP calf intestine alkaline phosphatase

CNV copy number variation

CTNNB1 β-catenin

ChIP chromatin immunoprecipitation

CpG cytosine-phospho guanine

DAPI 4',6-diamidino-2-phenylindole

DME differential miRNA expression

DME differential miRNA expression

DMEM Dulbecco'smodified Eagle's Medium

DMSO Dimethyl sulfoxide

DPH3 Diphthamide biosynthesis 3

ECM extracellular matrix

EDTA Ethylenediaminetetraacetic acid

EGF epidermal growth factor

ELF1 E74 Like ETS Transcription Factor 1

ERK extracellular signal-regulated kinase

Egr1 Early Growth Response

Egr1 Early growth response 1

EtBr Ethidium bromide

FCS fetal calf serum

FOLFOX folinic acid, 5-FU and oxaliplatin

GABPA GA Binding Protein Transcription Factor, Alpha

GEO Gene Expression Omnibus

GMCL1 germ cell-less, spermatogenesis associated 1

GRO-seq Global nuclear run-on sequencing

GSEA gene set enrichment analysis

GSEA gene set enrichment analysis

Gapdh Glyceraldehyde-3-phosphate dehydrogenase

Gusb β-glucuronidase

HBV hepatitis B virus

HCC hepatocellular carcinoma

HCV hepatitis C virus

HGF hepatocyte growth factor

HNRNPA1 Heterogeneous nuclear ribonucleoprotein A1

HSC hepatic stellate cells

HSP1A/B Heat Shock Protein Family A (Hsp70) Member 1A/B

IL interleukin

KSRP KH-Type Splicing Regulatory Protein

LANUV Landesamt für Umwelt und Naturschutz

LFU luciferase fluorescence units

LUAD lung adenocarcinoma

LUSC lung squamous cell carcinoma

Lamc1 laminin gamma 1

Loxl LOX-like

MAPK mitogen-activated protein kinase

MECP2 methyl-CpG binding protein 2

MMP metalloproteinases

MRTF-A myocardin-related transcription factor A

MYOD1 Myoblast determination protein 1

NAD^+ nicitinamide adenine dinucleotide

NADH nicitinamide adenine dinucleotide (NAD^+), reduced

NADPH Nicotinamide adenine dinucleotide phosphate, reduced

NASH non-alcoholic staetohepatatis

NF-$\varkappa$B nuclear factor κB

NFAT Nuclear factor of activated T cells

PBS Phosphate buffered saline

PCA Principal component analysis

PCR polymerase chain reaction

PDGFR platelet-derived growth factor receptor

PEG Polyethylene glycol

PES1 pescadillo ribosomal biogenesis factor 1

PIAF platinum, interferon, doxorubicin,5-FU

PPARγ proliferator-activated receptor γ

PVP Polyvinylpyrrolidone

PolII RNA polymerase II

RARs retinoic acid receptors (RARs)

RISC RNA-induced silencing complex

RNA-seq RNA sequencing

RNF185 ring finger protein 185

RPL13A ribosomal protein L13a

RPS20A ribosomal protein S20

RXRs retinoid x receptors

Rnu6 U6 small nuclear RNA

SAM S-adenosylmethionine

SDS Sodium dodecyl sulfate

SRF Serum response factor

Snord33 and small nucleolar RNA, C/D box 33

TACE transarterial chemoem-bolisation

TCA Citric Acid

TCF ternary complex factor

TCGA The Cancer Genome Atlas

TERT telomerase reverse-transcriptase

TGF-β transforming growth factor β

TIMPs tissue inhibitor of metalloproteinases

TNM Tumor,node and metastasis

TRBP TAR RNA-binding protein

TSS transcription start sites

Tbp TATA-box binding protein

Tris Tris(hydroxymethyl)aminomethane

UTR untranslated region

VDR vitamin D receptor

VEGF vascular endothelial growth factor

VEGFR vascular endothelial growth factor receptors

W-o-L Window-of-Linearity

ZEB1 Zinc finger E-box-binding homeobox 1

bp base pair

cDNA complementary DNA

dNTP Deoxynucleotide triphosphates

eEF-2 elongation factor 2

gDNA Genomic DNA

hTERT human telomerase

mRNA messenger RNA

miRNAs microRNAs

ncRNA non-coding RNAs

pHSCs primary HSCs

pre-miRNA precursor miRNA

pri-miRNA primary miRNA

qPCR quantitative PCR

sRNA-seq small RNA-sequencing

sRNA-seq small RNA-sequencing

Appendix C

Contributions

The following colleagues contributed to the main work presented in this thesis:

- Sebastian Winkler: analysed TCGA methylation and CNV data and performed linear regression analysis,
- Dieter Weichenhan and Marion Bähr: performed DNA methylation analysis of samples prepared by me,
- Jan Hengstler and Brigitte Begher-Tibbe: performed immunostaining and acquired images,
- Ralf Weiskirchen: isolated pHSC,
- Erawan Borkham-Kamphors: prepared CCl_4 RNA samples,
- Robert Geffers: prepared and sequenced sRNA-seq libraries,
- Abhishek Thavamani: isolated tumors from *SRF-VP16*^{iHep} mice and prepared RNA-seq library,
- Catrin Bitter: qPCR analysis of pHSC, CCl_4 and miRNA mimic transfected samples (performed jointly with me), cloning and mutagenesis of luciferase vectors, luciferase assays (performed jointly with me)

I performed the remaining work presented in the main part of the thesis.

Appendix D

Publications

Feliciello, I., Parazajder, J., **Akrap, I***., & Ugarković, Đ. First evidence of DNA methylation in insect Tribolium castaneum: environmental regulation of DNA methylation within heterochromatin. *Epigenetics*, **8(5)**, 534-541 (2013).

Feliciello, I., **Akrap, I***., Brajković, J., Zlatar, I., & Ugarković, Đ. Satellite DNA as a driver of population divergence in the red flour beetle Tribolium castaneum. *Genome biology and evolution*, **7(1)**, 228-239 (2014).

Feliciello, I., **Akrap, I.***, & Ugarković, Đ. . Satellite DNA modulates gene expression in the beetle Tribolium castaneum after heat stress. *PLoS genetics*, **11(8)** (2015).

Denisova, E., Heidenreich, B., Nagore, E., Rachakonda, P. S., Hosen, I., **Akrap, I***., ... & Sanmartin, O. Frequent DPH3 promoter mutations in skin cancers. *Oncotarget*, **6(34)** (2015).

Akrap, I.*, Thavamani, A., & Nordheim, A. Vps4A-mediated tumor suppression upon exosome modulation?. *Annals of translational medicine*, **4(9)** (2016).

Hermanns, C., Hampl, V., Holzer, K., Aigner, A., Penkava, J., Frank, N., ... **Akrap, I.**,... & Goppelt-Struebe, M. The novel MKL target gene myoferlin modulates expansion and senescence of hepatocellular carcinoma. *Oncogene*, **36(24)**, (2017)

Winkler, I., Bitter, C., Winkler, S., Weichenhan D., Thavamani, A., Hengstler J., ... & Nordheim A. Anti-fibrotic miRNA hubs modulated by Pparγ shape the fibrotic tumor microenvironment. Submitted

Winkler, I., Zhang, T., Rachakonda K., Klein, K., Eiermann, N., Brown, K., Kumar, R.

& Nordheim A. The influence of *DPH3* promoter mutation on ETS factor binding and gene expression. Manuscript in preparation

Winkler, I., Graeve, M., Thavamani A., Krämer A., Geffers R., & Nordheim A. Molecular characterisation of LT cell lines derived from hepatocellular carcinoma *SRF-VP16*^{iHep} mouse model. Manuscript in preparation

Winkler, I.$^{\#}$, Thavamani, A.$^{\#}$, Ducker, G.$^{\#}$, Matic, K., Weichenhan, D., Graeve, M., Orlich, M.,... & Nordheim A. Metabolic reprogramming of Serine Synthesis, Mitochondrial One-Carbon Metabolism, and Methionine Cycle Activity in Hepatocellular Carcinoma. Manuscript in preparation

Winkler, S., **Winkler, I.**, Figaschewski, M., Nordheim A., & Kohlbacher, O. DeRegNet - An exact algorithm for *de novo* identification of maximal deregulated subnetworks from multi-omics data. Manuscript in preparation

*maiden name
$^{\#}$equal contribution

The majority of the work presented here is contained in the manuscript Winkler et al.: Anti-fibrotic miRNA hubs modulated by Pparγ shape the fibrotic tumor microenvironment, submitted.

Work presented in the chapter **Summary of other projects**, subchapter **NonComs: Non-coding mutations in cancer** is contained in the manuscript in preparation Winkler et al.: The influence of *DPH3* promoter mutation on ETS factor binding and gene expression.

Work presented in the chapter **Summary of other projects**, subchapter **Characterization of the LT cell lines** is contained in the manuscript in preparation Winkler et al.: Molecular characterisation of LT cell lines derived from hepatocellular carcinoma *SRF-VP16*^{iHep} mouse model.

Work presented in the chapter **Summary of other projects**, subchapter **Metabolic reprogramming of serine synthesis, mitochondrial one-carbon metabolism, and methionine cycle activity in hepatocellular carcinoma** is contained in the manuscript in preparation Winkler et al.: Metabolic reprogramming of Serine Synthesis, Mitochondrial One-Carbon Metabolism, and Methionine Cycle Activity in Hepatocellular Carcinoma.

Appendix E

Supporting Figures

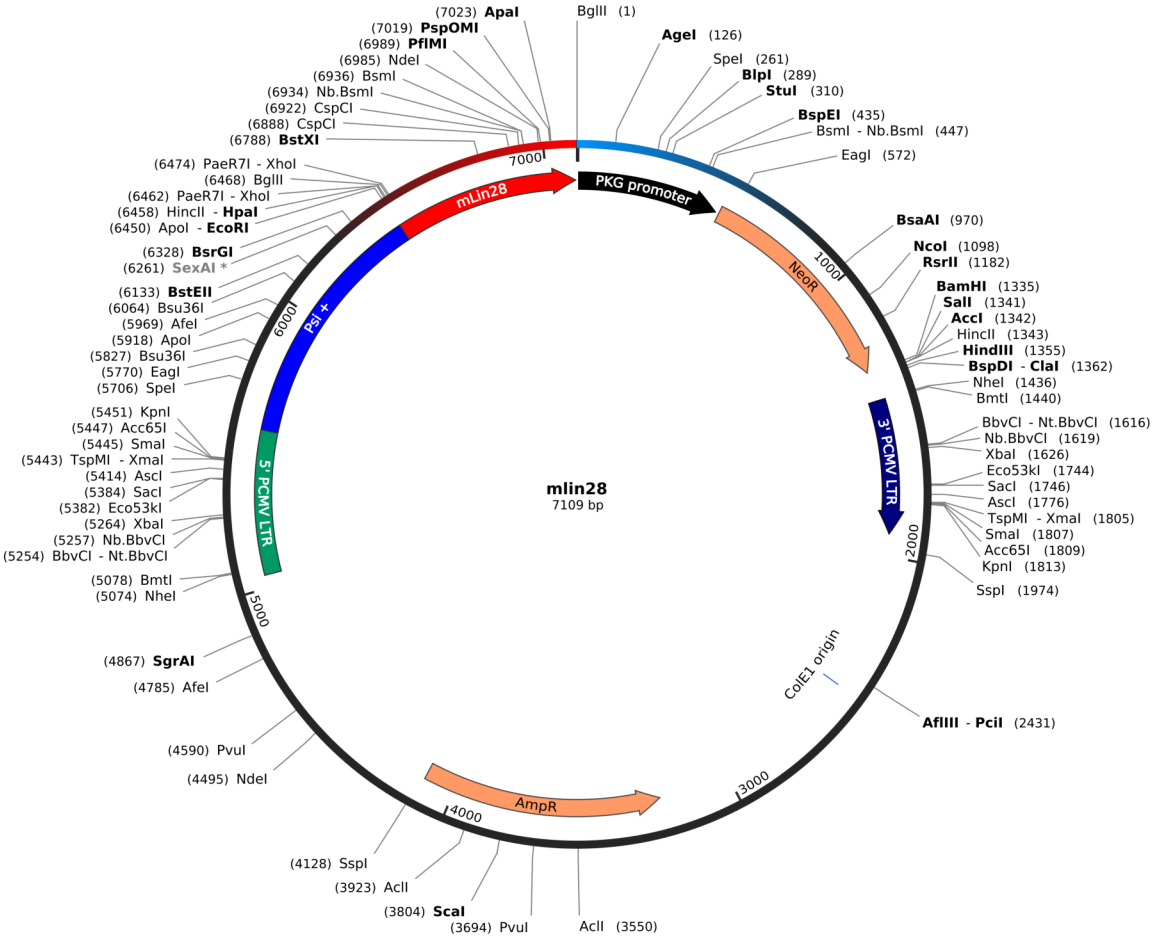

Figure E.1: Vector map of pMSCV-Lin28.

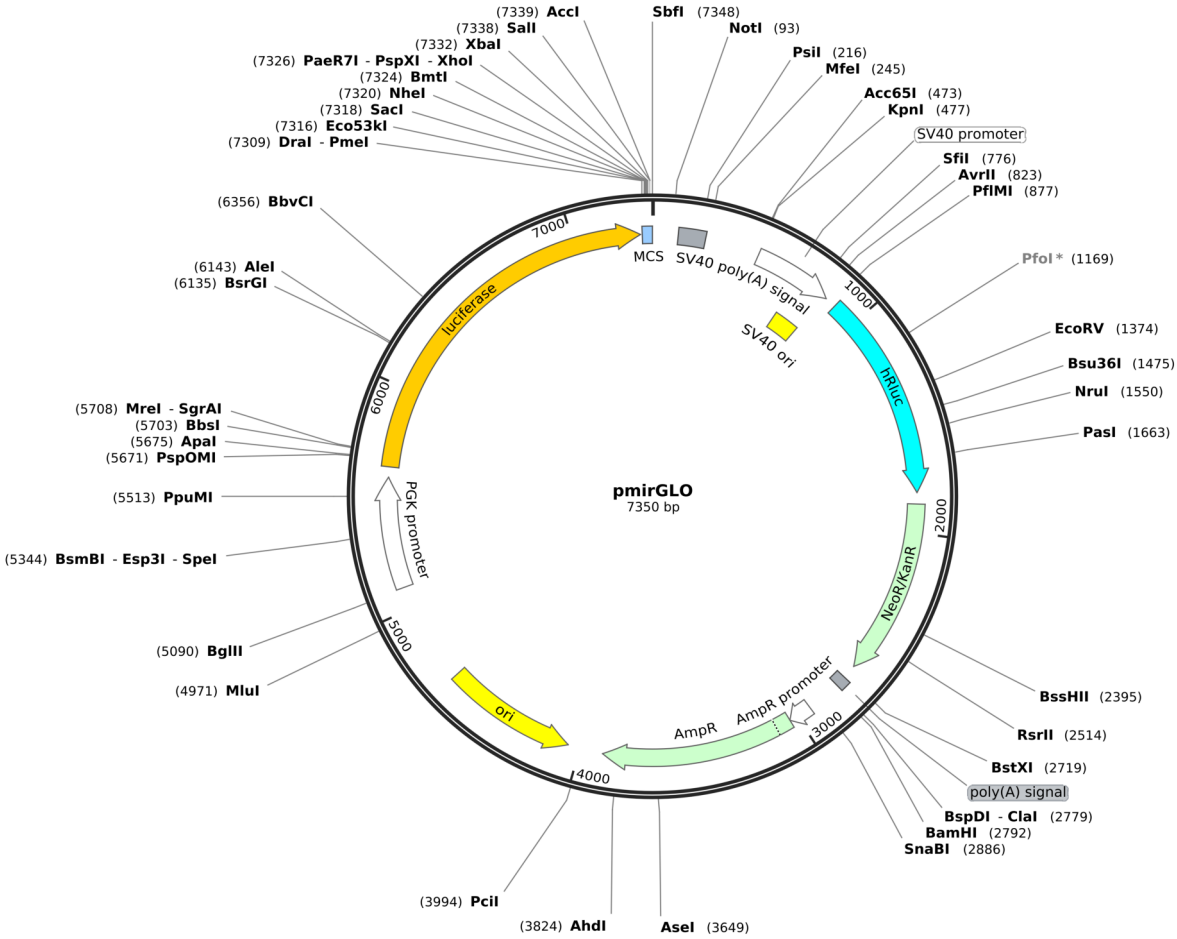

Figure E.2: Vector map of pmirGLO.

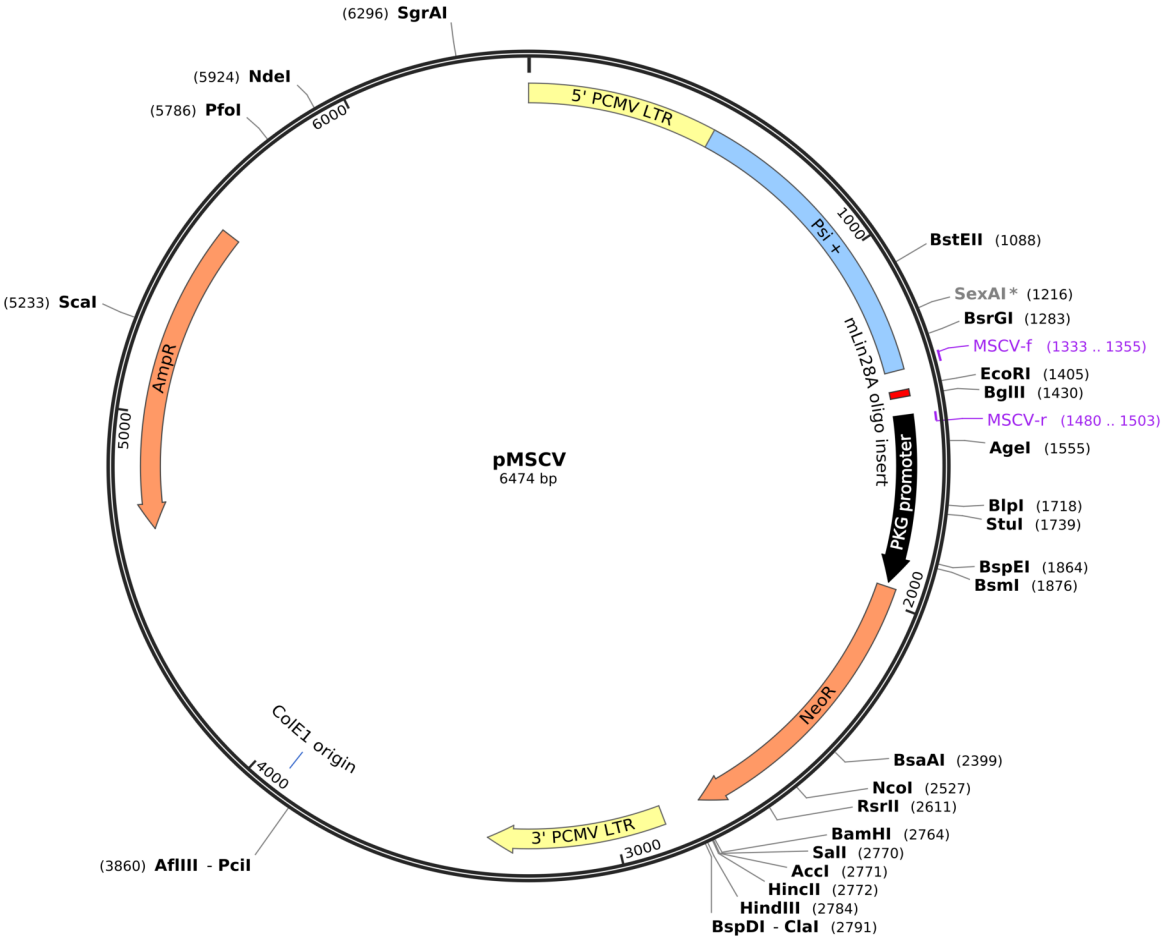

Figure E.3: Vector map of pMSCV.

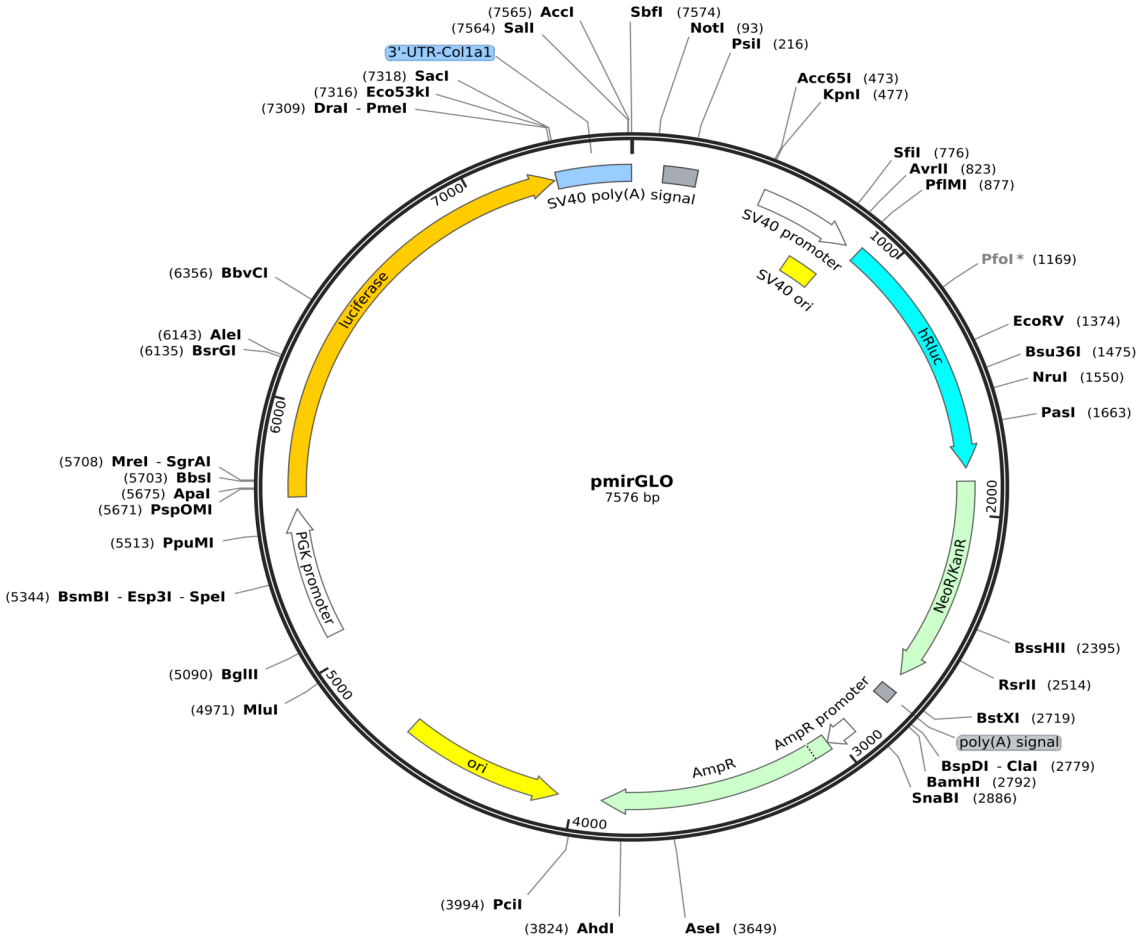

Figure E.4: Vector map of pmirGLO-Col1a1.

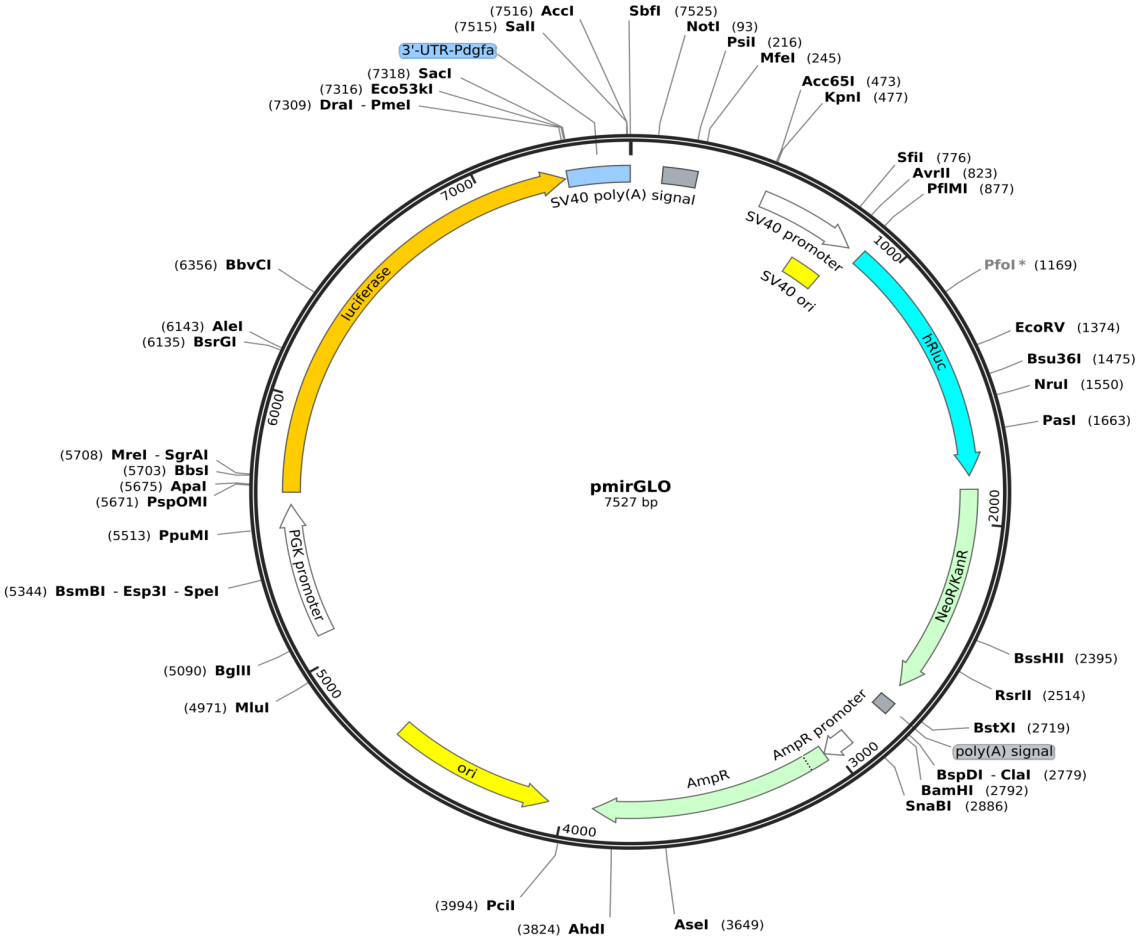

Figure E.5: Vector map of pmirGLO-Pdgfa.

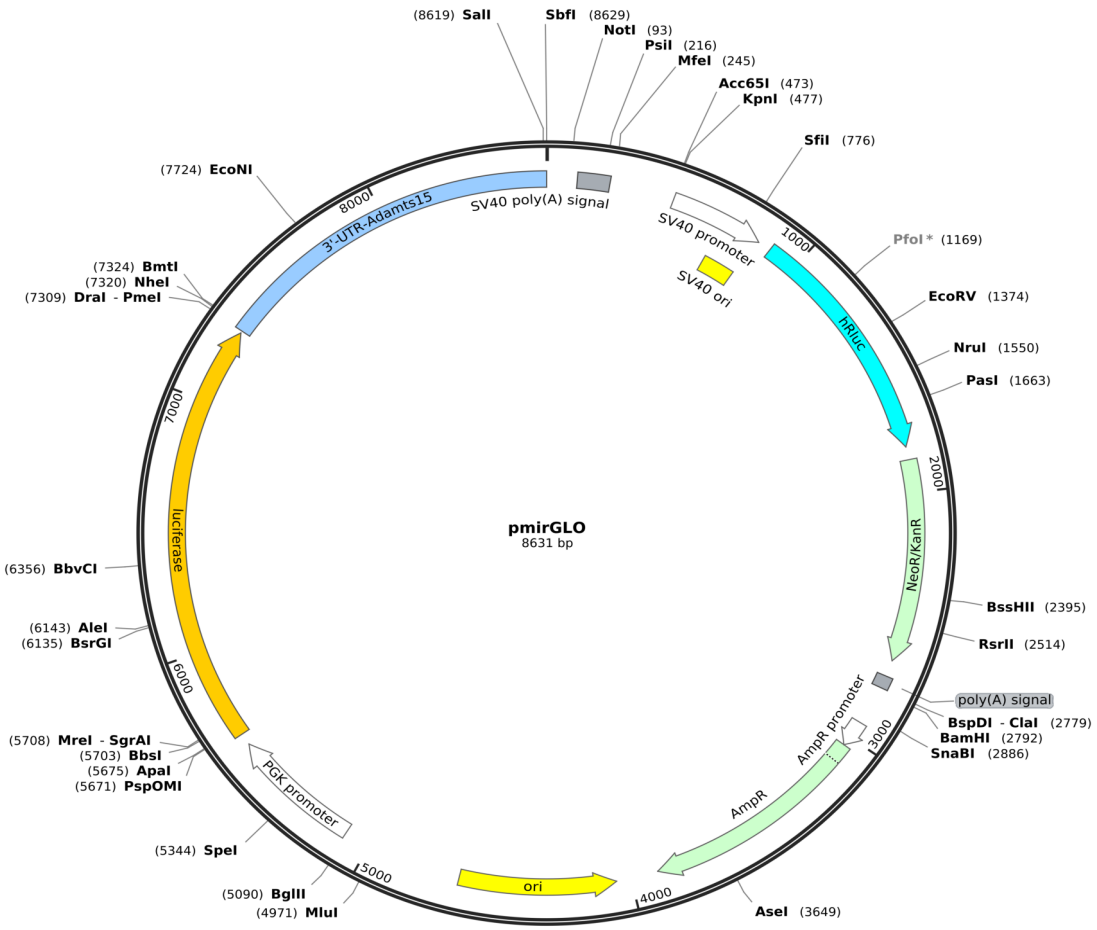

Figure E.6: Vector map of pmirGLO-Adamts15.

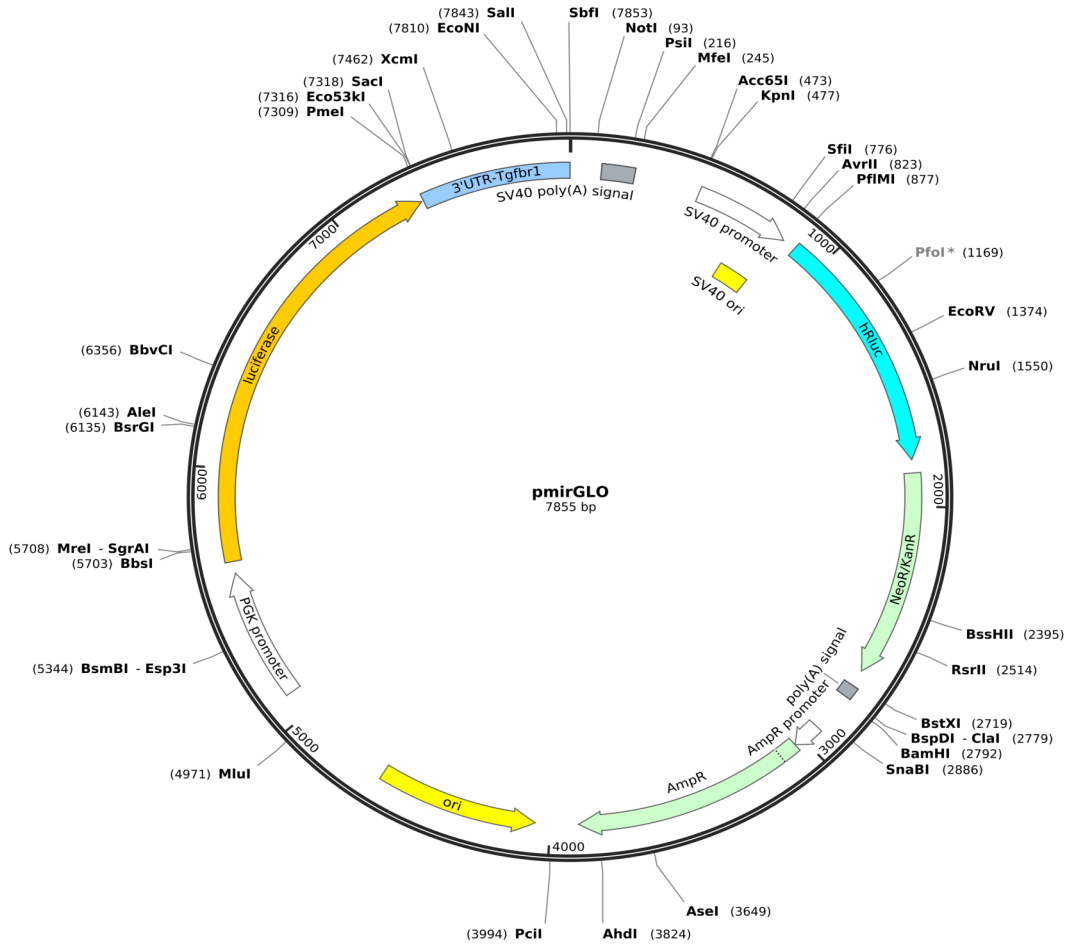

Figure E.7: Vector map of pmirGLO-Tgfbr1.

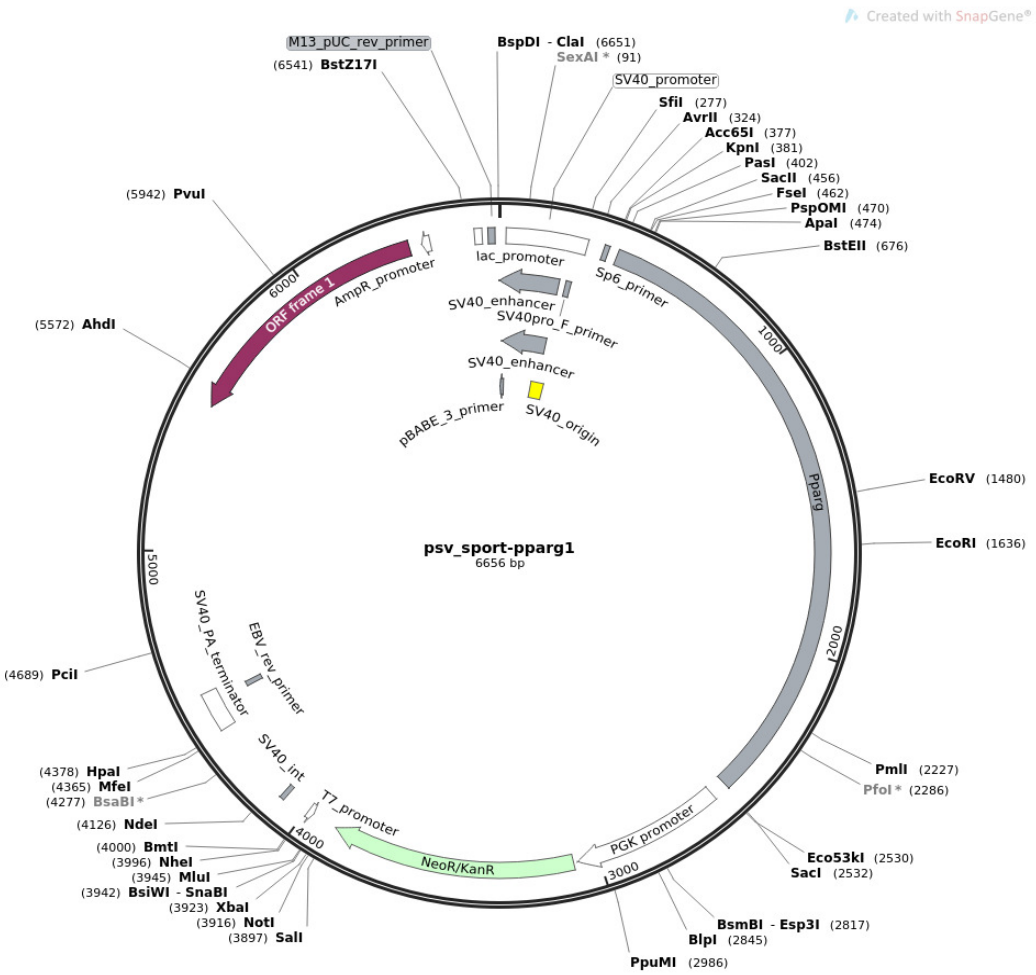

Figure E.8: Vector map of pSV Sport Ppar gamma 1.

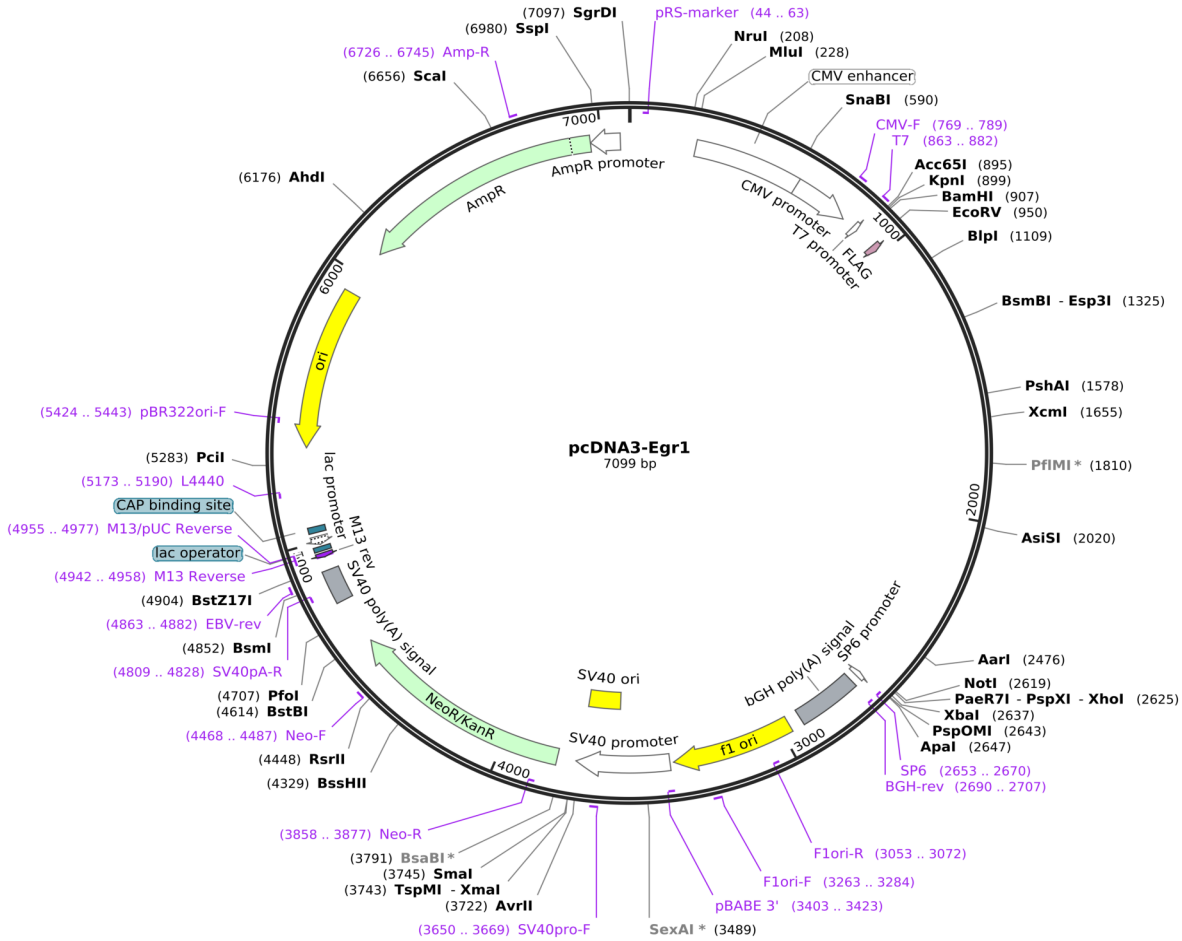

Figure E.9: Vector map of pcDNA3-Egr1.

Vector sequences and maps are provided in GENBANK format in CD attachment.

Appendix F

Supporting Tables

Suppl. Table 1. Summary of statistical analysis. Related to the figures as listed in the table. (See CD attachment)

Suppl. Table 2. List of differential expressed sRNAs and genes in SRF-VP16-driven nodular, tumor and corresponding control samples. (Sheet 1) List of differential expressed sRNAs obtained by sRNA-seq data analysis. List contains Gene ID information of all detected sRNAs, $\log_2$ fold changes in nodules and tumors compared to controls and corresponding p- and p_{adj}-values. (Sheet 2) Normalized read counts ($\log_2$ transformed) of sRNA in all samples. Read were produced using varianceStabilizing-Transformation function of DESeq2 package. (Sheet 3) List of differential expressed genes obtained by RNA-seq data analysis. List contains Gene ID information of all detected mRNAs, $\log_2$ fold changes in nodules and tumors compared to controls and corresponding p- and p_{adj}-values. (Sheet 4) Normalized read counts ($\log_2$ transformed) of mRNAs in all samples. Read were produced using varianceStabilizingTransformation function of DESeq2 package. (See CD attachment)

Suppl. Table 3. REC scores of linear regression analyses. (Sheet 1) REC scores of miRNA:mRNA pairs and the corresponding p_{adj}-values. (Sheet 2-5) Regression coef. of miRNA:mRNA pairs and the corresponding p_{adj}-values for miRNAs expression, CNV and methylation values in BRCA (Sheet 2), LIHC (Sheet 3), LUAD (Sheet 4) and LUSC (Sheet 5). (See CD attachment).

Suppl. Table 4. List of used primers, their sequence and targeted gene/miRNA. (Sheet 1) List of used primers, their sequence and targeted gene/miRNA. (Sheet 2)

Sizes and genomic coordinates of amplicons analyzed in miRNA methylation study. (See CD attachment).