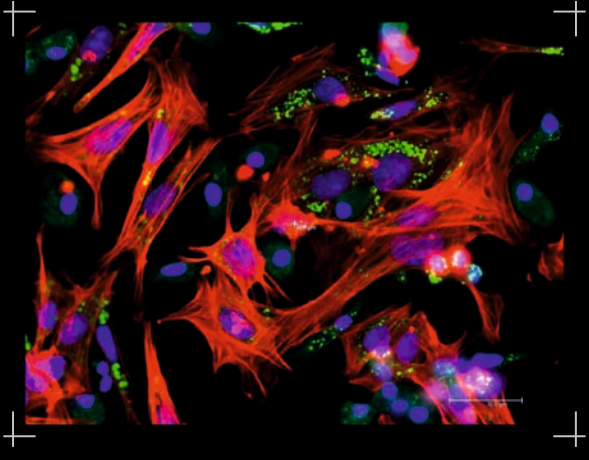
Identification of pulmonary PDGFRα-positive fibroblast specific *miRNA* and *mRNA* expression profiles during postnatal lung development

Daria Agnieszka Dontireddy



Inauguraldissertation zur Erlangung des Grades eines
Doktors der Humanbiologie
des Fachbereichs Medizin der Justus-Liebig-Universität Gießen

édition scientifique
VVB LAUFERSWEILER VERLAG

1. Auflage 2015

1st Edition 2015

Printed in Germany

édition scientifique
VVB LAUFERSWEILER VERLAG

STAUFENBERGRING 15, D-35396 GIESSEN
Tel: 0641-5599888 Fax: 0641-5599890
email: redaktion@doktorverlag.de

www.doktorverlag.de

Identification of pulmonary PDGFRα-positive fibroblast specific *miRNA* and *mRNA* expression profiles during postnatal lung development

INAUGURALDISSERTATION

zur Erlangung des Grades eines Doktors der Humanbiologie

des Fachbereichs Medizin

der Justus-Liebig-Universität Gießen

vorgelegt von

Daria Agnieszka Dontireddy (geb. Szerlowska)

aus Kędzierzyn-Koźle, Polen

Gießen 2014

From Max Planck Institute for Heart and Lung Research

Department of Lung Development and Remodelling

Bad Nauheim, Germany

Director: Prof. Werner Seeger, M.D.

First Supervisor and Committee Member: Prof. Werner Seeger, M.D.

Second Supervisor: PD Dr. Robert Voswinckel, M.D.

Committee Members: Prof. Dr. Wilhelm Wößmann, M.D., Prof. Dr. Dr. Hans-Rudolf Tinneberg, M.D., Prof. Dr. Birgit Lorenz, M.D.

Date of Doctoral Defense: February 27th 2015

Table of contents

Chapter 1

INTRODUCTION

1.1. Stages of lung development

The mammalian lung development can be divided into several pre- and postnatal phases (Figure 1) [29, 56]. In rodents it starts around 9th gestational day and in humans at the 4th gestational week (embryonic stage). During this stage the trachea is formed from the foregut lung bud and separated from the esophagus. The distal part of the tracheal bud starts to form two primary bronchial buds of the future right and left main bronchi. The Pseudoglandular stage starts around 12th gestational day in rodents and around the 8th gestational week in human. During this time the bronchi develop into secondary (lobar bronchi) and tertiary bronchi (segmental bronchi). Along with the branching of the airways, endodermal- and mesodermal-derived cells start to appear [56]. These endodermal derived cells include neuroendocrine cells, basal cells, and ciliated/secretory cells (clara/goblet cells) [92]. Whereas mesodermal derived cells include: smooth muscle cells, fibroblasts and cells forming cartilage of the bigger airways [29, 92]. The next canalicular stage starts about 16.5th gestational day in rodents and at the 16th gestational week in humans. The terminal bronchioles start to form respiratory bronchioles and alveolar ducts lined by nonciliated epithelial cells [56]. Vascularization and capillarization also begins at this stage [29, 92]. The most important respiratory system structure establishes about 5th postnatal day in rodents and at the 36th gestational week in humans during the terminal saccular stage [56]. At this stage the primitive alveoli are formed. The respiratory epithelium and the endothelium as well as fibroblasts interact to give rise to the morphological area of the gas exchange region. At the same time alveolar epithelial type II cells starts to secret pulmonary surfactant, which reduces the lung surface tension [29, 56]. In rodents, the alveolar stage takes place after birth, whereas in humans, fully functional alveoli start to develop before birth. During alveolarization, the terminal saccules and alveolar ducts increase in number and the saccular walls are divided by primary septa. The newly formed

alveoli are the most important units of gas exchange in mammals [19]. The stage of alveolarization lasts up to postnatal day 30 in rodents and up to two years of age in humans. However further secondary septa formation may still be continued [19, 56, 92, 106]. Since the objective of this current work is to understand the late lung development processes of alveolarization and septation, these processes are discussed in detail in the next section.

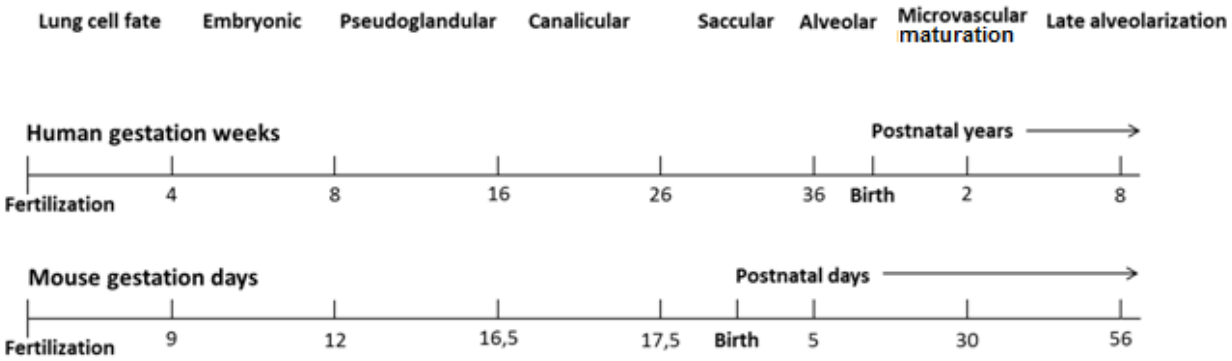


Figure 1 Comparison of human and mouse lung development stages timing. *Arrows indicate the duration of the process of alveolarization. Modified according to [56].*

1.2. Alveolarization

The central layer of connective tissue "sandwiched" between two capillary networks (primary septum) represents the basic structure necessary for alveolarization [19]. During the process of alveolarization, the gas-exchange area is enlarged by subdivision of the terminal sacs (septation). The outgrowth of the secondary septum from the primary septum leads to the formation of new interalveolar walls. Elastic fibers of the connective tissue produced by fibroblasts and the reorganization of the capillary network contribute to these key structural changes. A large number of new bulges elongate and form secondary septa with the double capillary network outlining the future alveoli (Figure 2) [19, 103]. This process of "bulk alveolarization" is terminated at about 2 weeks after birth in rodents and 12–24 months in humans. The last stage of lung development, the microvascular maturation starts at postnatal day 14 (P14) and lasts until P30 in rodents. In humans it starts several months after birth and may last till 2-3 years, even 8 years of age [19]. At this stage the alveolar septum double capillary layer is transformed into a single-layered capillary

network [15, 19, 103] and the absolute mass of the intercalated septal interstitium is reduced [19, 98, 103].

Gas exchange occurs across the mature alveolar septum between the finest barriers of type I epithelial cells, the basement membrane (specialized form of extracellular matrix) and endothelial cells of the capillaries [4, 19]. The anatomical structures of gas exchange are so-called alveoli, which are lined by respiratory epithelium and wrapped in a fine mesh of capillaries covering about 70% of its area. Septation process depends on mechanical forces produced by elastin and collagen fibers. These fibers form extracellular matrix (ECM) and are produced by mesenchymal fibroblasts. Fibroblasts carry out certain function during alveolar septation and their dysfunction might be a cause of many lung diseases [4].

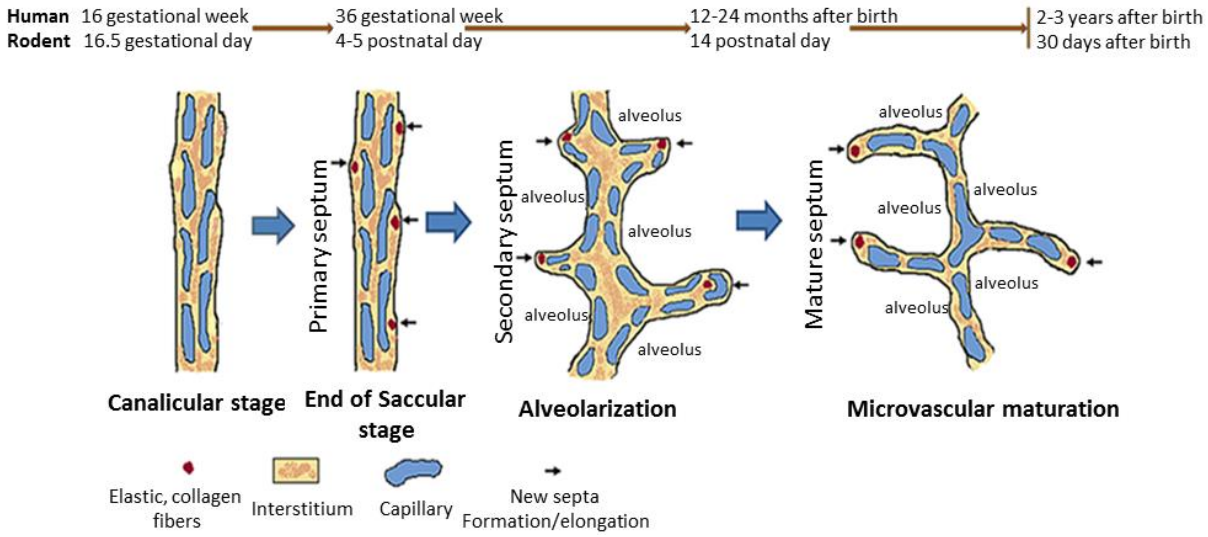


Figure 2 Formation of new septa and new alveoli. *Accumulated at the tip of new septa elastin and collagen fibers produce forces that push the tissue into the saccular airspace where alveolar sacs are subdivided into definitive alveoli (secondary septation). The immature septa (primary and secondary) contain double capillary network, which becomes single capillary network during the microvascular maturation, moreover the interstitium (connective tissue) thinning occurs.*

1.3. Lung interstitial fibroblasts

Previous multiple works have evidenced that there are three distinguished heterogeneous fibroblast subsets present in lung [67, 79, 81, 94]. These fibroblasts subsets are:

- Lipofibroblasts,
- Myofibroblasts and
- Platelet-derived growth factor receptor alpha (PDGFRα)-positive cells.

It is known that the fibroblast subpopulations are able to transdifferentiate to each other [57]. Each of this fibroblast subset plays an essential role during lung development and pulmonary diseases. They can be characterized by specific cytoskeletal or surface markers as well as a spatial and temporal appearance pattern in lung tissue.

1.3.1. The lipofibroblast

Morphology and phenotype:

In the immature rodent lung, the lipid-containing interstitial cells consist for about 50% of the resident alveolar wall cells [94]. Lipofibroblasts are interstitial mesenchymal cells which are characterized mainly by lipid droplets content. These cells were described in the early 1970s in the fetal and neonatal rat lung [88]. In early 1980s they were called lipid interstitial cells (or lipid droplet–laden), ipso facto distinguishing them from non-lipid interstitial cells (NLIC). Lipofibroblasts have characteristics similar to other mesenchymal cells, such as adipocytes, pericytes or smooth muscle cells. However, they contain more glycogen and are localized to the central region of the alveolar septum. Lipofibroblasts contain intermediate contractile filaments similar to those observed in contractile myofibroblasts or contractile interstitial cells (CIC). In contrast to the NLICs, lipofibroblasts are Thy-1-positive (Thy-1^{+}) [81]. Thy-1 is a glycosylphosphatidylinositol-linked cell-surface glycoprotein discovered on T lymphocyte surface and involved in transmembrane signal transduction. Thy-1 plays a role in lipofibroblast differentiation and also influences the regulation of lipid homeostasis via peroxisome proliferator-activated receptor-γ (PPARγ) activity [81, 119]. PPARγ is a nuclear transcription factor which seems

to play an essential role during fibroblast differentiation into lipogenic phenotype and prevents interstitial fibroblast to myofibroblast differentiation [94]. Thy-1^+ and Thy-1^- subsets of lung fibroblast differ morphologically when cultured in vitro. Thy-1^+ fibroblasts are more elongated, spindle-shaped, and have filopodia which suggest their contractile function, whereas the Thy-1^- cells are more round, spread and lack filopodia [81, 119]. On the other hand only Thy-1^-, but not Thy-1^+ human orbital fibroblasts differentiate into lipofibroblasts when stimulated with a PPARγ agonist, such as prostaglandin J_2 [58].

In rodents, the lipofibroblasts are detectable in a large amount at gestational day 16. In the second neonatal week, their number decreases due to apoptosis and reduced proliferation [81, 98]. Lipofibroblasts are also observed in adult rats and mice lungs, although they contain much less lipids than in the neonate. In the neonate rat lung approximately 85% of the lipid-laden cells constitute of lipofibroblasts, whereas the remaining 15% constitutes of macrophages. Lipofibroblasts and the non-lipid interstitial cells exhibit similar ECM and cytoskeletal proteins profile in cell culture [81].

Function:

It has been shown that lipofibroblasts are the essential cells assisting pulmonary surfactant production and epithelial cell differentiation into alveolar type II cells. Thus they play a critical role during normal lung development, homeostasis, and later adult lung injury and repair [81, 94]. Lipofibroblasts provide high levels of triglycerides to the neighboring type II cells which are necessary for phospholipid synthesis (main component of surfactant) in their surfactant-secretory granules, so-called lamellar bodies [81]. The highest accumulation of triglycerides in rodent pulmonary fibroblasts occurs just before birth. Lipofibroblast and alveolar type II cell interaction is facilitated by a specific molecule, the parathyroid hormone-related protein (PTHrP). It is secreted by alveolar type II cells and subsequently binds to the PTHrP receptor on lipofibroblasts. PTHrP/ PTHrP receptor signaling pathway stimulates lipofibroblast differentiation by down-regulating the Wnt signaling pathway. Moreover leptin which is secreted by lipofibroblasts binds to its receptor on the alveolar type II cells and facilitates the mesenchymal-epithelial cross talk [113]. Epithelial-mesenchymal interaction up-regulates PPARγ and its downstream target adipose differentiation-related protein (ADRP). The ADRP is a lipid storage droplet-

coating protein that facilitates triglyceride uptake by lipofibroblast and alveolar type II cells [74, 94, 104, 119].

Lipofibroblasts are also the source for retinol and retinoic acid (RA), the biologically active forms of retinoids outside the retina. RA is involved in the regulation of gene expression during early postnatal lung development. Lipofibroblast nuclear receptor, the PPAR-γ forms heterodimer with the retinoid-X receptor (RXR), and binds to responsive elements in the regulatory regions of target genes. The heterodimer might be accelerated by prostaglandins secreted by alveolar type II cells [81].

1.3.2. The myofibroblast

Myofibroblasts are a differentiated type of interstitial fibroblasts responsible for wound healing and organs remodeling. They fulfill two major functions in the lung: extracellular matrix (ECM) synthesis and fibers-related tension production [67]. Boström, et al. showed that lung alveolar myofibroblasts are critical for the formation of secondary septa, and their absence is associated with a lack of functional alveoli [14].

Morphology and phenotype:

In normal conditions pulmonary myofibroblasts exhibit phenotype with few actin-associated cell-cell and cell-matrix interactions and a little ECM production. Activated myofibroblasts migrate and acquire contractile phenotype. The key factors responsible for fibroblasts differentiation are cytokines and growth factors released by inflammatory cells and residing epithelial cells. Furthermore, changes in the mechanical microenvironment (eg. extracellular stress) of the lung can also influence myofibroblast differentiation [47, 48]. Differentiated myofibroblasts are characterized by *de novo* expression of α-smooth muscle actin (α-SMA). α-SMA represents the stress fiber which is connected with the ECM at the sites of supermature focal adhesions and with cells via adherent junctions. α-SMA is encoded by the *Acta* gene and its regulation is complex and depends on a variety of factors. Pro-fibrotic cytokines, transcription factors, mechanical forces and stiff ECM with its specialized proteins like the ED-A splice variant of fibronectin are involved in *Acta* gene regulation. α-SMA is activated by TGFβ, whose signaling is modulated by SMAD3 [47,

48]. Platelet-derived growth factor (PDGF) initiates chemoattraction and differentiation of these cells. Fibroblast growth factor (FGF) directly orchestrates cell proliferation, whereas indirectly effects the alpha and beta transforming growth factors (TGFα, TGFβ) [67].

Lung alveolar myofibroblasts are characterized by the expression of intermediate filaments, i.e. vimentin which stabilizes cellular architecture, α-SMA, non-muscle myosin (NNM) and smooth muscle myosin heavy chain isoform 1 (SM-MHC or SM1). Moreover their phenotype and ultrastructure resemble proximal airway smooth muscle cells and fibroblasts. SM-MHC seems to be the marker for fully differentiated myofibroblasts as this protein expression is up-regulated later than α-SMA during myofibroblasts differentiation [67]. Unlike vimentin, desmin is expressed mostly in myofibroblasts at the alveolar ducts but not in those at the tip of the alveolar septum. Additionally, a family of actin-binding proteins, the drebrin is transiently expressed in myofibroblasts providing the mechanical elongation of secondary septa [127]. During lung alveolarization, myofibroblasts are the major cells found within the septal tips and are associated with elastin deposition [67].

Function:

Myofibroblasts are responsible for tissue contractility or compliance determined by their stress fibers. Stress fibers are the structures in non-muscle cells which consist of actin filaments, crosslinking proteins, and myosin II motors. Myofibroblasts are the main source of ECM whose components (elastin and types of collagen) provide elastic properties to the lung parenchyma and maintain alveolar integrity for normal respiration and lung function. Elastin and collagen fibers form deposits in the thick walls of terminal air sacs and are elongated by the mechanical forces that give rise to primary and secondary septa [16, 48]. Thus myofibroblasts are the main cells driving alveolarization. Their number increases after birth and decreases after the microvascular maturation. In the adult lung, alveolar myofibroblasts are almost no detectable [48, 127]. However their number and proliferation capacity has been shown to increase during lung injury or fibrotic disease. During normal lung development and healing process, myofibroblasts progression and proliferation is terminated by apoptosis or their dedifferentiation [48].

Uncontrolled myofibroblasts proliferation and functions are the main cause of interstitial pulmonary fibrosis (IPF). In this state they produce excessive amount of

collagen and extracellular matrix deposition. Fibrotic lesions formed within thickened alveolar walls decreases the ability of gas exchange between the epithelium and the endothelium, and reduces lung compliance [67].

1.3.3. PDGFRα-positive cell

Platelet-derived growth factor (PDGF) and its receptor (PDGFR):

The PDGF molecule was first time isolated from human platelets. It is a covalent dimer of two different or same polypeptide chains A, B, C or D. The chains are linked to each other by disulfide bonds, encoded by distinct genes and inactive in their monomeric forms. There are four known homodimers and one heterodimer for PDGF molecules: PDGF-AA, -BB, -AB, -CC, and -DD [9]. Activated PDGFs more or less specifically bind to three different PDGF receptor forms: PDGFRαα (PDGFRα), PDGFRββ (PDGFRβ) or PDGFRαβ. The receptors also form active non-covalent dimers held together by the bivalent PDGF ligands. Heterodimeric PDGFRαβ delivers the highest range of signals than any of the other two homodimers. The extracellular ligand-binding sites of these receptors consist of five immunoglobulin-like domains (each of the chain). PDGFR-αα specifically interacts with PDGF-AA, PDGF-BB, PDGF-CC and PDGF-AB; PDGFR-ββ binds with high affinity to PDGF-DD and PDGF-BB; PDGFR-αβ is PDGF-AB and PDGF-BB specific (Figure 3). Some of the PDGF/ PDGFR interactions can be forced *in vitro* in cell culture. For example, PDGF-BB may activate PDGFRα, whereas this interaction seems to be not essential for PDGFRα-mediated signaling *in vivo* [9].

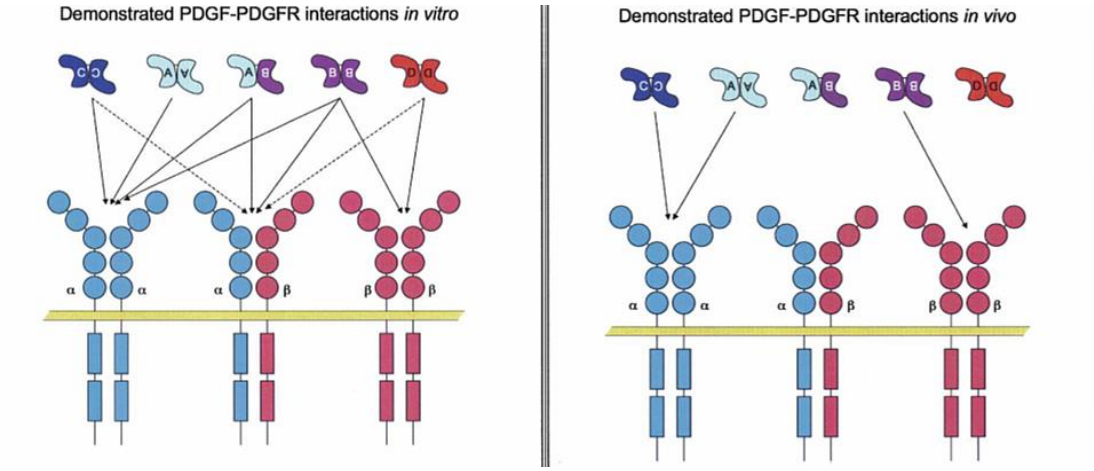


Figure 3 Members of tyrosine kinase receptor family (RTK), the PDGFRs and their activation patterns by different isoforms of PDGF. *Extracellular IgG domains and cytosolic homodimeric or heterodimeric RTK domains are shown. Known in vitro and in vivo interactions are presented. Adopted from [7].*

The PDGFRα (CD140a) is a ~122 kDa transmembrane tyrosine kinase receptor (RTK). It is enzymatically activated following ligand binding and auto-phosphorylation at multiple sites in the cytoplasmic domain of the molecule. In the dimeric receptor, the two subunits act as substrates for each other's tyrosine kinase. Their phosphorylation increase the kinase activity and form docking sites for SH2 (Src Homology 2) domain-containing downstream signaling molecules. PDGFRα and PDGFRβ signaling activates several signal transduction pathways as the key downstream mediators, i.e.: Ras/mitogen-activated protein kinase (MAPK), phosphatidylinositol 3-phospahte kinase (PI3K), and phospholipase-γ (PLCγ). These signaling pathways contribute to the regulation of different cellular processes such as cell proliferation, differentiation, survival, cell motility and immunoresponse [9].

PDGFRs and their ligands are organ and cell types specifically expressed. In the lung PDGF-A, PDGF-B, PDGF-C, PDGFRα, and PDGFRβ are expressed. Moreover, PDGF-B is expressed in vascular endothelial cells and appears to signal via PDGFRβ which is expressed in vascular smooth muscle cells (vSMC); PDGF-A and -C are expressed mostly in the epithelium and bind to the PDGFRα which is expressed by the interstitial fibroblasts. These two paracrine signaling loops do not interfere with each other. The PDGFs secretion and activation are compartmentalized in the extracellular space and its bioavailability is tightly regulated during development [9, 10, 49, 68].

PDGFRα-positive cells morphology and phenotype:

During early lung development, namely the pseudoglandular stage, PDGF-A is expressed by the lung epithelium, whereas the PDGFRα-positive cells are mainly located in the mesenchyme surrounding the bronchial endbuds. During the canalicular and saccular stages, PDGFRα-positive cells spread along the surface of the future terminal sacculi (alveolar sacs) walls and differentiate into alveolar myofibroblasts. These cells differentiation is essential for alveolarization. Parabronchial smooth muscle cells (PBSM)

also express PDGFRα, however on much lower level. This suggests that alveolar myofibroblasts and smooth muscle cells may share the same cell lineage [14, 71].

Function of PDGFRα-positive cells:

In 1985 Kaplan, et al. proposed that the lipofibroblasts are the precursors of the myofibroblasts [53]. However more likely, the precursor cells of myofibroblasts are considered as platelet-derived growth factor receptor α (PDGFRα)-positive mesenchymal cells [14]. PDGFRα is not only a marker of myofibroblast precursor, but activated by its ligand (simultaneously mitogen for smooth muscle cells, fibroblasts, and glia cells) has also functional consequences [71]. It promotes proliferation of PDGFRα-positive progenitor cells and regulates their differentiation, migration and survival [13, 14, 71]. PDGF-A/ PDGFRα signaling is required for postnatal alveolarization (septa elongation) but not for early lung development when the lung branching occurs [13].

During late lung development the PDGF-A/ PDGFRα interactions are tightly related through epithelial-mesenchymal cross-talk. PDGF-A knockout in rodents causes a failure of alveolar septum formation. Homozygous PDGF-A-null mice are born with defects during alveolarization process, i.e. complete loss of myofibroblasts and septal elastin deposits [14, 71], while the PDGFRα-null mice die at the early embryonic stage or after birth from lung dysfunction [110]. Furthermore, overexpression of PDGF-A in the lung epithelium causes overgrowth of the lung mesenchyme, resulting in compression of the distal respiratory airways and lethality [68]. Until now no other genes have been shown to regulate alveolar myofibroblast differentiation.

PDGFRα expression level positively correlates with the proliferation and differentiation capability of fibroblasts. It has been shown, that cells with higher PDGFRα expression (GFP expression in nuclei as marker of endogenous *Pdgfrα* gene) demonstrate increased αSMA protein expression level and proliferation during the septa formation in rodent [54]. Moreover, it has been shown that PDGF-A/ PDGFRα signaling inhibits lipofibroblast-promoting pathways [70] or the formation of adipose tissue [2]. On the other hand, neutral lipids co-localize with PDGFRα-GFP$^{bright+}$ cells which reside in the proximity of alveolar type II cells but not in the PDGFRα-GFP^{dim+} cells of the bronchiolar wall of the adult mouse lung [5].

1.4. MicroRNA

MicroRNAs (miRNAs) are ~22nt long small endogenous non-coding RNAs that influence post-transcriptional gene regulation (RNAi – RNA interference). One miRNA can interact with a wide range of different target genes as well as many different miRNAs can synergistically influence one target gene [1, 6, 72]. The complexity and diversity of miRNAs play a key regulatory role in cell homeostasis and development, and their dysfunctions are implicated in many diseases [72].

1.4.1. Discovery

In 1993 the group of V. Ambros discovered a lin-4 gene in heterochronic mutants of *Caenorhabditis elegans*. It did not encode a protein but contained a small segment of homology to multiple motifs in the 3'-untranslated region (3'-UTR) of another heterochronic protein-coding gene lin-14 [66]. The discovery of another *C.elegans* miRNA gene let-7 with broad conservation among metazoans [91] helped to understand the importance of these small RNAs and initiated a rapid increase of studies on RNA interference biochemistry and miRNA biogenesis. Present miRNA studies involve both biological and bioinformatic methods. All known miRNAs are compiled in an official database miRBase (http://www.mirbase.org/) together with their hairpin transcripts, sequences and gene locations. The latest miRBase v20 contains 24521 hairpin precursor miRNAs expressing 30424 mature miRNA products in 206 species [59, 72]. Many other computational methods using specific algorithms were developed to identify putative miRNA targets (also including empirically validated targets) or biological processes in which miRNAs are involved [63, 72, 90, 107].

MiRNA nomenclature: capitalized "miR"-X refers to the mature miRNAs; miRNAs with nearly identical sequences except for one or two nucleotides differ with small letter miR-Xa, miR-Xb, etc.; pre-miRNAs which lead to 100% identical mature miRNAs but that are located at different places in the genome: mir-X-1,-mir-X-2; the species of origin is designated with a three-letter prefix, e.g., hsa-miR-X (*Homo sapiens*), mmu-miR-X (*Mus musculus*), 'v' for viral and 'd' for *Drosophila*; when two mature

microRNAs originate from opposite arms of the same pre-miRNA, they are denoted with a -3p or -5p suffix; when the relative expression levels are known, an asterisk indicates miRNA expressed at low levels relative to the miRNA in the opposite arm of a hairpin, miR-X and miR-X* share a pre-miRNA hairpin. X refers to any miRNA number.

1.4.2. Biogenesis

MiRNAs are endogenously expressed small RNAs that undergo extensive processing and are part of the RNA-induced silencing complex (RISC). In animals miRNAs are mostly transcribed as separate coding genes (60%; canonical biogenesis). About 15% of them are polycistronic genes coding multiple miRNAs. 25% of miRNAs are coded by short introns of protein-coding genes called mirtons (alternative biogenesis) [27, 34, 122]. Polycistrons which contain more than two miRNA genes form miRNA clusters which members are not necessarily identical but may have similar expression patterns and target several categories of genes (physical relatedness). Homologous miRNAs that share similar sequence are collected in miRNA families [40]. The first step of miRNA biogenesis starts when the precise RNase III family enzymes liberate miRNA from its precursor transcripts (Figure 4). The primary miRNA (pri-miRNA) is transcribed by DNA-dependent RNA polymerase II (RNAPII) and similarly to protein-coding genes has 5'-end cap structure and 3'-end polyA tail [27, 34, 122]. The miRNA transcription is regulated by many transcription factors frequently in tissues-specific manner [61]. Pri-miRNA contains a region of imperfect dsRNA known as the stem–loop structure (hairpin) that is enzymatically removed later. The cleavage process begins when a double strand RNA binding domain (dsRBD) protein complex, the Pasha/DiGeorge syndrome critical region gene 8 (DGCR8) binds to the pri-miRNA simultaneously recruiting the Drosha enzyme (RNase III). This multiprotein complex is called the microprocessor. Drosha cleavage relieves a ~60–70nt precursor miRNA (pre-miRNA) with 2-nt single-stranded 3′ overhangs. The 3′ overhangs are recognized by the nuclear export protein Exportin 5 which actively transports pre-miRNA in a Ran–GTP-dependent manner to the cytoplasm. In the cytoplasm, pre-miRNA is cleaved by Dicer and forms a ~22nt miRNA:miRNA* duplex.

Dicer leaves its signature on the both ends of small miRNA duplexes, the 2-nt single-stranded 3′ overhangs [27, 34, 122].

Mirtrons do not undergo the conventional biogenesis. They overpass Drosha processing using the splicing machinery to generate pre-miRNAs. Excised, debranched and refolded they form short stem–loop structures that mimic pre-miRNAs and are processed into mature miRNAs by Dicer [27].

The mechanism of small RNA sorting that guides a particular small RNA strand to load into a specific Argonaute (AGO) family member is not entirely understood, however this is extremely crucial for the biological function of miRNAs. Argonaute proteins perform as a core of the RNA-induced silencing complex (RISC) and determine miRNAs biological function. Once a miRNA duplex is made, it seems that one strand is assessed and its fate determined. Only one miRNA-duplex strand is selectively stabilized during RISC assembly. Predominantly, it is the evolutionary preferred guide strand (miR strand) which binds complementary to its targets. The second strand – passenger strand (miR* strand) is discarded and degraded [27]. The strands arise from independent precursor molecules and different types of AGO proteins may compete for the selection of these strands from each duplex. This shows that miR* are not non-functional molecules, like thought before [27, 128]. The proper strand choice depends on the enzymatic dicing of pre-miRNA that may influence the intrinsic structure, terminal nucleotides or thermodynamic properties of the miRNA duplex strands [27]. MiRNA sorting is also determined by structural interaction with AGO domains. The PAZ domain commonly hosts the 3’-OH termini of the miRNA, whereas the Mid domain forms a binding pocket that anchors a 5’-phosphate of the terminal nucleotide. The AGO C-terminal PIWI domain shows similarity to RNase H fold and harbors Asp–Asp–His residues responsible for catalytic activity (endonucleolytic cleavage of target gene transcripts) [27, 34]. In mammals, only AGO2 (mouse Eif2c2) forms an active endonuclease, although the catalytic triad is conserved in all four members of the AGO protein family (AGO1, AGO2, AGO3, AGO4) [21, 30, 34].

The maturation of the miRNA/RISC complex is not well understood. One of the proposed mechanism is that miR* strands dissociate in a cleavage-independent manner by unwinding – a process that is facilitated by the presence of mismatches in the loaded duplexes [27, 51].

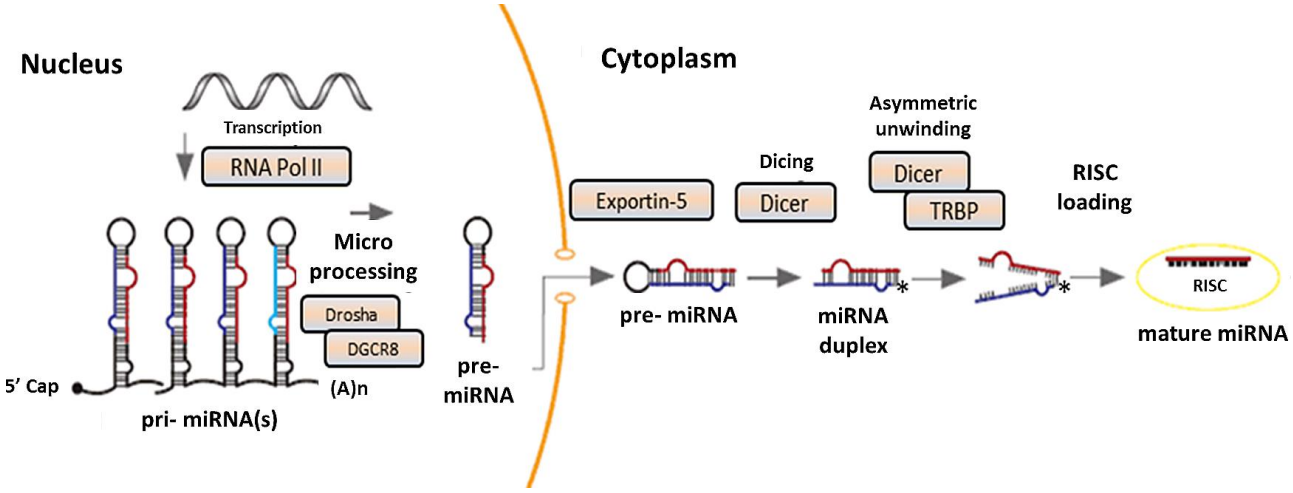


Figure 4 Schematic summary of miRNA biogenesis. *Boxes show enzymes or other proteins which are crucial for miRNA biogenesis and "maturation". TRBP – TAR RNA binding protein, it recruits the Dicer complex to AGO2. * passenger miRNA strand. Modified according to [37, 122].*

1.4.3. Mechanism of action

MiRNAs regulate post-transcriptional gene expression. This regulation is accomplished by mature RNA-induced silencing complex (miRISC) formed by the guide miRNA strand and protein complex. The miRNA acts to guide the RISC complex to its target mRNA, while the AGO protein complex represses mRNA translation or induces deadenylation-dependent mRNA decay [27]. Most of the miRNAs bind to the 3'-UTRs (untranslated regions) of the target transcripts, less commonly to open reading frame (ORF) and 5'-UTRs. There are two patterns of binding (Figure 5). First, when the target site has perfect Watson–Crick complementarity to the 5'-end of the miRNA "seed region". This nearly perfect complementarity is sufficient for miRNAs to suppress their targets or mRNA cleavage without requiring further base pairings at the 3'-end of the miRNAs [130]. MiRNA-directed target cleavage has only been reported in a few cases, and it is assumed to be the principal regulatory mode for endo-siRNAs and for piRNA-mediated repression [27]. The second type of binding has imperfect complementary base pairing at the 5'-end of the miRNAs and it is compensated via additional base pairings in the 3'-end of the miRNAs (3' compensatory binding). This type of binding causes translation repression [130].

Activity of miRISC is optimized and regulated by many factors. These factors may bind to mRNA targets facilitating/stabilizing or counteracting miRISC activity. It is also known that an auto-regulatory feedback loop exists between the miRNAs and their target genes. This is greatly important during the cell fate determination and the development of organs. For instance, one miRNA can repress a mRNA that encodes factors involved in its own biogenesis or function, whereas the same mRNA may inhibit or activate this miRNA [61].

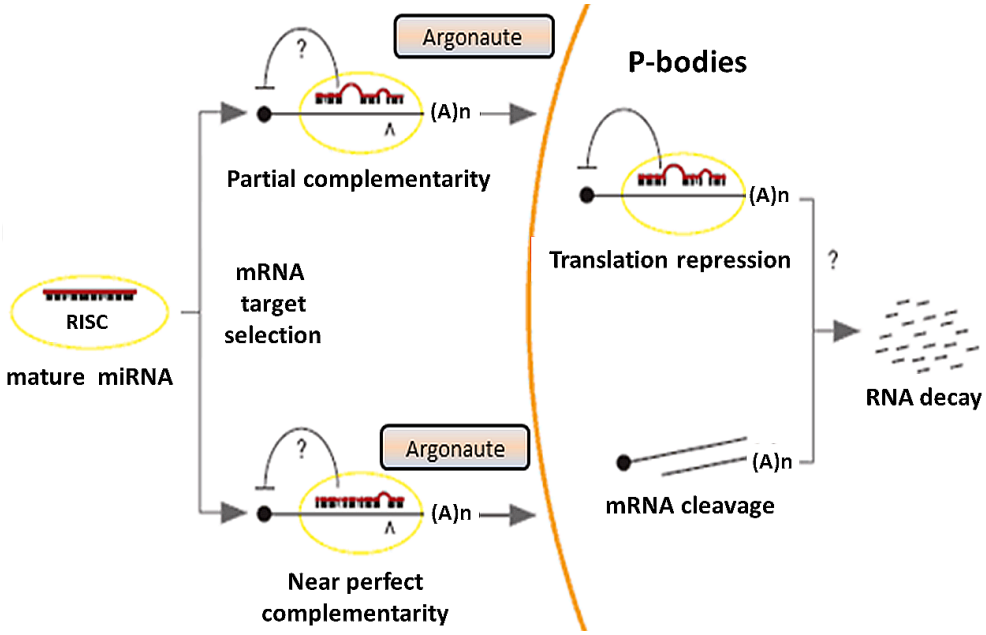


Figure 5 The mechanism of action of miRNAs. *Two different pathways of mRNA targets fate are shown (described above). mRNA processing takes place in processing bodies (P-bodies), the distinct foci within the cytoplasm of the cell consisting of many enzymes involved in mRNA turnover. Modified according to [37, 122].*

1.4.4. MiRNAs in lung diseases and therapeutics

MiRNAs are known to be key players involved in mammalian organs development and pathogenesis. It has been shown that in the lung miRNAs have a very specific expression profile highly conserved among the mammalian species. During lung development, miRNA networks are dynamically regulated [31]. The knowledge of the miRNA functions in physiological and pathological lung conditions is still poorly

understood. However, there is a known group of miRNAs involved in lung inflammation, viral infections, tumorigenesis and pulmonary diseases (Table 1) [112]. Many of these diseases are associated with lung structure remodeling due to uncontrolled mesenchymal or epithelial cells proliferation or apoptosis. In turn these cellular dysfunctions might be caused by many factors including up- or down-regulation of specific miRNAs or their target genes. In this case therapeutic approaches would require replacement of specific miRNA duplexes or reducing a miRNA's level by using modified antisense oligonucleotides, so-called antimiRs or antagomirs [101]. Alternatively, gene therapy techniques could be used, i.e. DNA vectors that encodes a therapeutic protein (overexpression strategy) [82] or antisense siRNA/shRNA against dysfunctional mRNA (knockdown strategy) [121].

Table 1 MiRNAs involved in lung development, homeostasis and deregulated in pulmonary diseases [38, 83, 84, 105, 112, 117, 129].

miRNA in the lung	
Homeostasis and lung development (fibroblasts regulation)	miR-155, miR-26a, let-7, miR-29, miR-15/miR-16, miR-223, miR-146a/b, -145, miR-17-92 cluster
inflammation and viral infections	miR-146a/b, miR-21, -25, -27b, -100, -140, -142-3p, -181c, -187, -194, -214, -223 and -224, miR-200a, miR-155, miR-17-92 cluster, miR-574-5p
Immune-mediated lung diseases	miR-148a/b, miR-152, miR-21, miR-126, let-7, miR-29a, miR-155, miR-133a
Cancer	miR-155, let-7, miR-17-92 cluster, miR-21, miR-210, miR-218, and miR-34 family
Chronic obstructive pulmonary disease (COPD)/ emphysema	let-7 family, miR-15b, -34a, -199a-5p, -125b, -144, -145, -146a, -150, -1274a, -424, -107, -101, -452, -449, -203, -222, -340, -223, -18a, -106a, -99a, -365
Idiopathic pulmonary fibrosis	let-7d, miR-21, -145, -29, -154, -155, -338*, -127
Cystic fibrosis	miR-126, -101, -494, -138, -155

Pulmonary artery hypertension	miR-17-92 cluster, miR-150, -21, -204, -27a, -17, -210, -206, -145, -143, -126
Acute lung injury (ALI) and acute respiratory distress syndrome (ARDS)	miR-21, -55, let-7, miR-146, -32*, -466d-5p, -466f-3p, -127, -16

Chapter 2

HYPOTHESIS and AIMS

During late lung development, the stabilization and differentiation of the phenotype of fibroblasts subtypes are regulated in a time-dependent manner by specific molecular mechanisms. Dysfunctional fibroblast subtypes represent a central cause of structural lung diseases, i.e. pulmonary fibrosis, emphysema or COPD. PDGFRα-positive fibroblasts have been shown to be involved in the physiological process of alveolarization since they serve as precursor cells for myofibroblasts. Loss of the PDGFRα system leads to a complete failure of septum formation. While during lung development many key factors decide about the cell fate and hence whole organ, the same key factors might be involved in the lung disease onset. Therefore the understanding of the fibroblasts subsets function and differentiation biology is needed to develop new therapeutic regenerative strategies. Based on this background, the aim of the present study was to investigate mouse pulmonary PDGFRα-positive cells at different developmental time points and characterize these cells for:

A) phenotype and differentiation capacity,
B) miRNome and
C) transcriptome

in order to identify new possible target molecules.

Chapter 3

METHODICAL BASIS for miRNome and TRANSCRIPTOM ANALYSIS

Molecular biology techniques were used to screen the whole miRNome (miRNA profile) and transcriptome (gene expression level) of PDGFRα-positive cells. TaqMan Low Density Array (TLDA) was used to analyze the miRNome, whereas Next-Generation Sequencing (NGS) was used to analyze the transcriptome.

3.1. TaqMan Low Density Array

The following method is based on TaqMan qPCR (Figure 7) and it requires the use of specific primers that can transcribe small miRNAs into cDNA. These small RNA-specific stem-loop primers are specific for the 3'polyadenylated ends of miRNA (Figure 6).

Figure 6 Reverse transcription (RT) of miRNA. *Small RNA-specific stem-loop RT primers are specific for the 3'polyadenylated end of microRNA. Adopted from handbook part no. 4425446 Rev. A, 1/2009 (Applied Biosystems).*

TaqMan miRNA Arrays are specific for all mature miRNAs and can distinguish between miRNAs that differ by only one nucleotide.

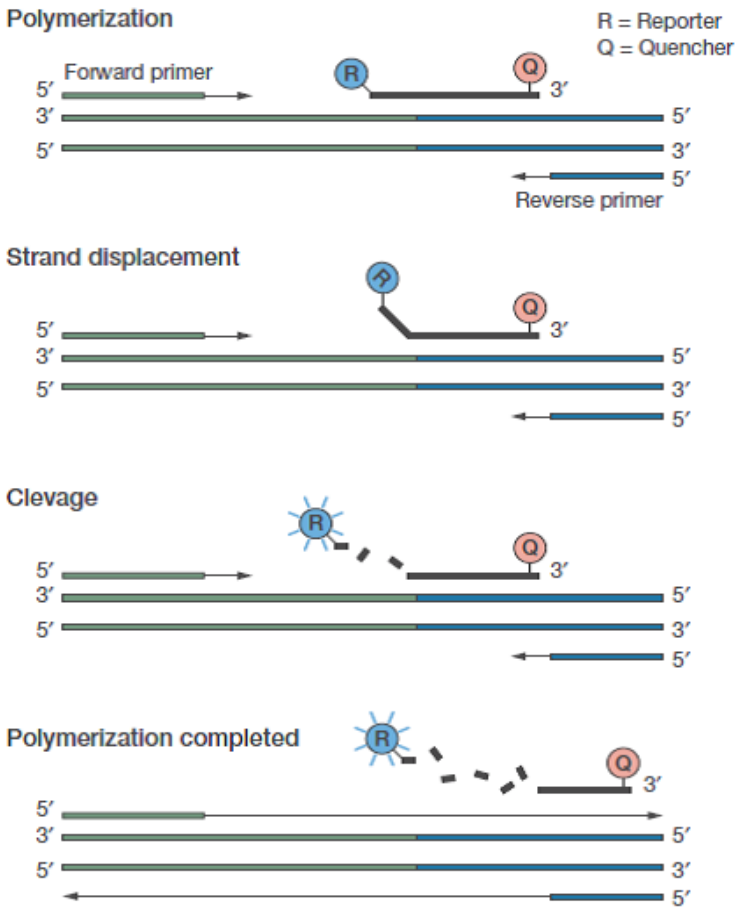


Figure 7 Steps in TaqMan real-time PCR. *Each TaqMan minor groove binder (MGB) probe contains: a reporter dye (FAM™ dye) linked to the 5'-end of the probe, MGB at the 3'-end of the probe [nonfluorescent quencher (NFQ)]. The short probes have great specificity to the target cDNA and increase the sensitivity of qPCR. The 5' to 3' nuclease activity of the DNA polymerase system allows amplifying cDNA. Adopted from handbook part no. 4425446 Rev. A, 1/2009 (Applied Biosystems).*

3.2. Next-Generation Sequencing

In 1977 Fred Sanger and Alan R. Coulson published two methodological papers on rapid determination of DNA sequence [dideoxynucleotide (ddNTP) chain termination sequencing]. Sanger sequencing relies on enzymatic synthesis of DNA in the presence of chain-terminating inhibitors of DNA polymerase, i.e. ddATP, ddCTP, ddGTP and ddTTP. These four modified nucleotides are characterized by the absence of a hydroxyl group and

inhibit DNA strand elongation. In Sanger sequencing analysis, DNA synthesis reaction is carried out four times in parallel i.e. with only one type of dideoxynucleotide and DNA products are identified by radioactivity [radioactive deoxynucleotide (dNTP) 32P- or 35S-dATP] on polyacrylamide gel [99]. However, use of advanced fluorescent dye-labelled ddNTPs permits sequencing in a single reaction with all four ddNTPs (use of capillary gel electrophoresis and CCD camera with signals translated into a chromatogram) [109]. Sanger sequencing method marked the beginning of a new era for deciphering complete genes and later entire genomes [99]. In the last few years this sequencing technique has been replaced by Next-Generation Sequencing (high-throughput sequencing) methods, especially for large-scale automated genome or transcriptome analyses. DNA cloning necessity was eliminated by DNA immobilization and repeated amplification. High-throughput sequencing technologies are intended to lower the cost of DNA sequencing by generating a massive number of short sequence reads (short sequence of nucleotides) in a single experiment (thousands or millions of sequences concurrently) [41, 52]. The next generation sequence reads are produced from fragmented DNA libraries which later must be assembled together. For RNA sequencing, the libraries are produced from fragmented RNA molecules that are transcribed into cDNA. Relatively little input DNA/ RNA is needed to produce a library. The DNA or RNA are fragmented, tagged by specific adaptor oligos and amplified. The majority of next-generation sequencers produce shorter read lengths (35–250 bp) than capillary sequencers (650–800 bp); what also can impact the utility of data for various applications such as *de novo* assembly. As no sequencing technology is perfect and each instrument will generate different types and amounts of errors, it is necessary to understand, identify and exclude error-types that may impact the interpretation of downstream analysis [76].

Different next-generation sequencing methods are commercially available:

a) Single-molecule real-time sequencing (Pacific Bio)
b) Sequencing by synthesis
    i. Ion Torrent – detection by ion release
    ii. Illumina – detection by fluorescent dyes
    iii. Pyrosequencing (454)
c) Sequencing by ligation (SOLiD sequencing)

In the current work we aim to sequence the RNA isolated from pulmonary PDGFRα-positive cells using Ion Torrent sequencer.

RNA-Seq (sequencing), so-called Whole Transcriptome Shotgun Sequencing (WTSS) uses the capabilities of next-generation sequencing to reveal a snapshot of RNA presence (gene presence or alternatively spliced transcript of a gene) and quantity (expression level of genes) at any particular time in a specific cell type, tissue or organ. The number of mRNA molecules encoding an individual gene provides information about specific cell phenotype (when counting protein-coding genes) [23].

3.2.1. Ion Torrent semiconductor sequencing

It is a system based on standard sequencing chemistry (A, C, G, T) directly translated into digital information (0, 1) on a semiconductor chip. The method is based on detection of hydrogen ions (H^+) released during the DNA polymerization. A micro-well containing a single-stranded template DNA (library) is flooded with a single type of nucleotide. If the introduced nucleotide is complementary to the leading template nucleotide, it is incorporated into the growing complementary strand. This causes the release of a hydrogen ion that triggers a hypersensitive ion sensor (by changing pH), which indicates that a reaction had occurred. Homopolymer repeats i.e. multiple nucleotides of the same type are incorporated in a single cycle. This leads to the release of the corresponding number of hydrogens and a proportionally higher electronic signal [97].

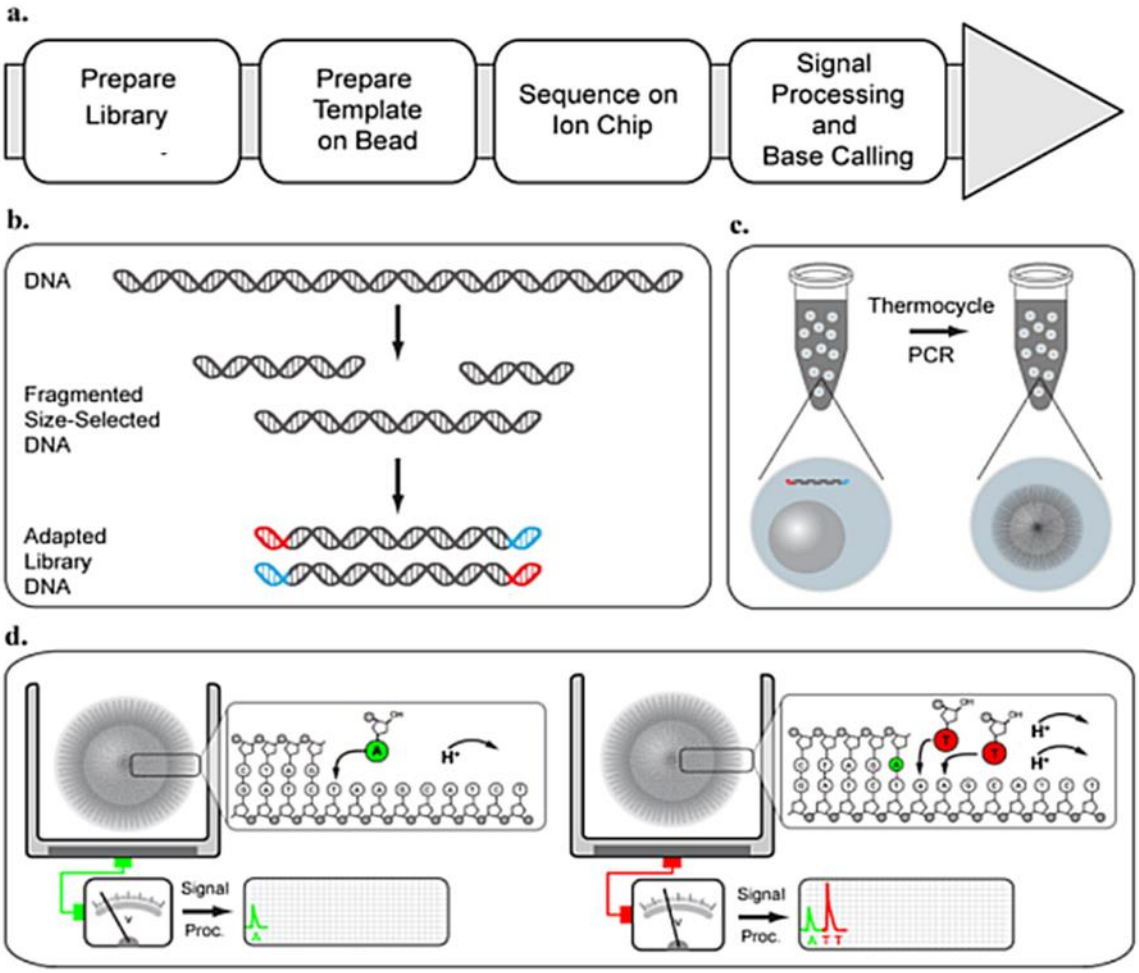


Figure 8 Steps of the Ion Torrent sequencing process (a). *DNA library fragmentation and ligation to forward/reverse adapters (before this step RNA must be reverse transcribed into cDNA) (b). Clonal amplification of the DNA fragments occurs on the template-carrying beads during emulsion PCR (c). Polymerization of nucleotides releases H^+ ions causing a pH change. The detection occurs at the thin layer (semiconductor) located at the bottom of each micro-well. If homopolymer repeats (e.g. AAA) are present, the electronic signal is higher. All four nucleotides are cyclically flowed in an automated run (d). According to [96].*

3.2.1.1. Preparation of cDNA libraries

Total RNA used for transcriptome sequencing needs to be purified from the cytoplasmic and mitochondrial ribosomal RNA (rRNA) which constitutes the majority (>90%) of the total RNA. It can be achieved by RNA hybridization to bait oligos bound to

beads. Alternatively, poly(A) selection of mRNA can be used for this purpose. rRNA removal increases the capacity to retrieve data from the remaining portion of the transcriptome (mRNA), and small non-coding RNAs. The RNA has to be fragmented by RNase III treatment and then reverse transcribed into complementary DNA (cDNA). The purpose of preparing a library is to link the DNA fragment at its both ends to forward/reverse adapters with bar codes (barcoded primers). Similar distribution between the samples is desired; therefore median size of the library fragments is selected (100, 200, 300 or 400 bases). For sequencing of RNA other than mRNA, the library preparation is modified. For small RNA targets, such as miRNA, RNA is isolated through size selection. The library molar concentration and average size of the fragments are monitored using a bioanalyzer chip. This information determines the dilution factor which is necessary to set the ideal ratio of DNA fragment library/ Ion Sphere Particle (ISPs) for the next stage of emulsion PCR [12, 25].

3.2.1.2. Preparation of the sequencing template

Preparation of the sequencing template is an automated step that allows the clonal amplification on the surface of ion sphere particles (beads) to which a fragment library is amplified. The clonal amplification is performed during the emulsion PCR (ePCR) (Figure 9) and helps to achieve a detection threshold signal necessary during the process of sequencing. Three parts enable the automated delivery of the templated ion sphere particles. First, a specific microenvironment is necessary to create millions of microreactors in which clonal amplification occurs. Secondly, a fully integrated thermal cycler and a disposable path amplification plate system are needed. Thirdly, an integrated centrifuge recovers the templated ion sphere particles. The biotinylated PCR primers allow the enrichment of the templated ion sphere particles from the reaction mix by binding to streptavidin-linked magnetic beads. Monoclonality of the amplified cDNA is desired because it alone is the source of the sequencing data, i.e. only ISPs with a single fragment from the library could be analyzed [12, 25].

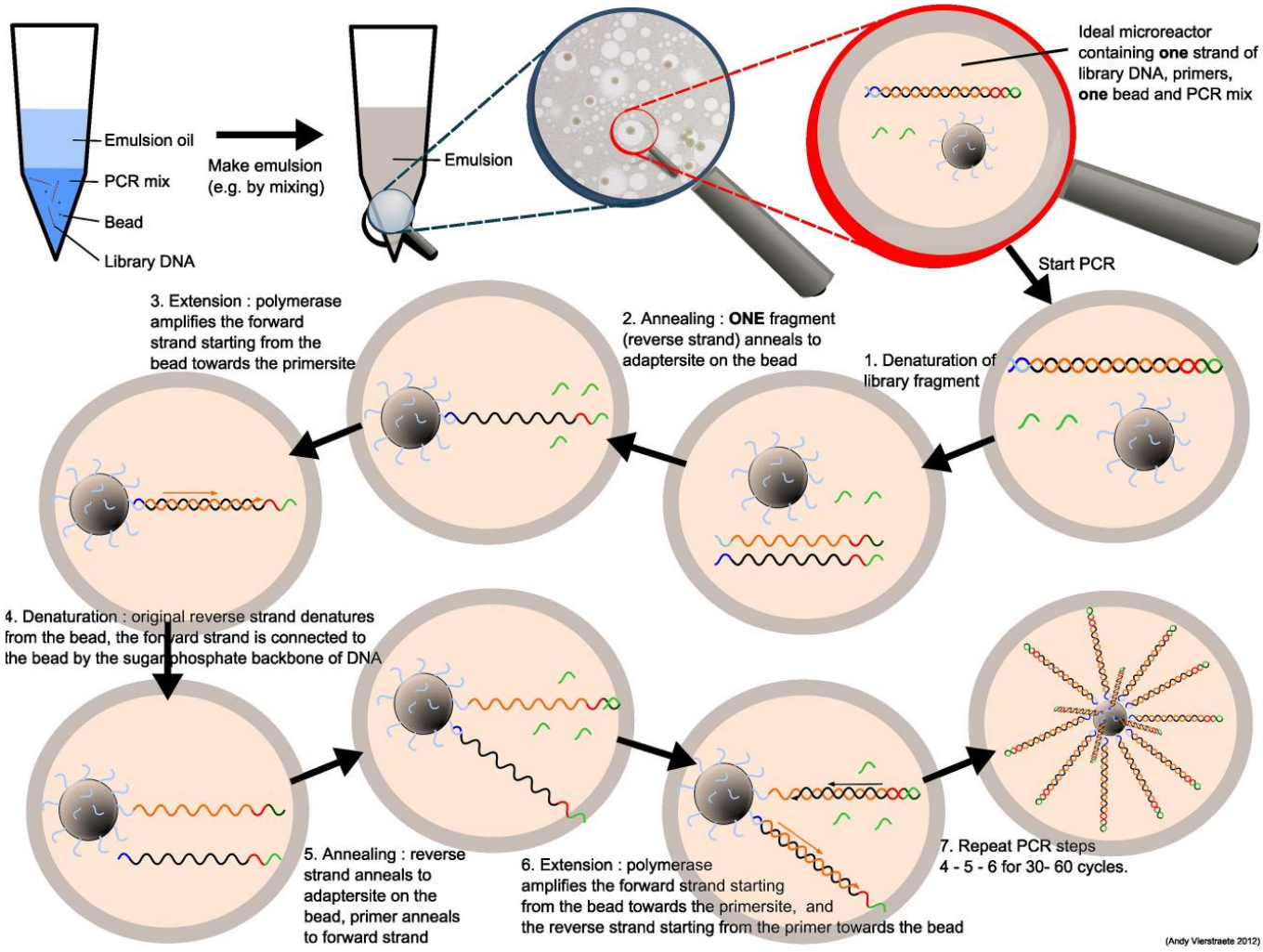


Figure 9 Emulsion PCR steps. *According to Andy Vierstraete. Department of Biology, Faculty of Science, Ghent University. 2012.*

3.2.1.3. Sequencing

Prior to sequencing, all reagents are booted to a pH ~ 7.8. The sequencing template, the primers and the polymerase are loaded onto the chip. For example, Ion PI™ Chip v2 may provide: 10 giga base pairs (Gb) of sequence data in 60-80 million reads. The adapters contain sequences that allow the primers to bind and initiate the sequencing reads. Each polymerization of unmodified nucleotides releases H^+ ions causing a pH change. Detection occurs at the thin layer (semiconductor) located at the bottom of each micro-well [12, 25].

3.2.1.4. Data quality and depth of coverage

All raw data is generated as ionograms and transferred to the Torrent Server. The algorithm "base calling" allows data (reads) conversion into FASTQ format i.e. text-based format of nucleotide sequence letters A, G, C, T and quality score. In FASTQ format each nucleotide is associated to an ASCII (American Standard Code for Information Interchange)-encoded quality number corresponding to a Phred score (Q) which is directly translated into the probability (p-value) of incorrect base call [24, 28]. These quality scores are generated by a Phred-like algorithm (FastQC software and other trimming tools) to determine highly accurate consensus sequences and to estimate their quality (the probability that a base is called correctly). The scores are listed as Q values where Q20 score is considered as an acceptably accurate base call (>99% accuracy; Q30>99.9%, etc.). Data trimming serves a dual purpose. First, it monitors parts of the reads with poor quality and secondly, it removes potential adapters' contamination. Trimming of the sequencing data gives a better result in the further analysis. Automated DNA sequencers occasionally produce poor quality reads, particularly near the sequencing primer site and towards the end of longer sequence runs [12]. Additional filters are used to remove wildly varying lengths or polyclonal reads i.e. missequenced when two different templates are amplified on a single bead, resulting in a hybrid sequence [12, 100]. Depth of coverage stands for the average number of times a given region (sequence of nucleotides) has been sequenced by independent reads. Deep sequencing indicates that the depth of the process is many times larger than the length of the sequence under study (low expression of a gene can be detected). Correlation of gene abundance between the samples or replicates can be measured as a Fragments Per Kilobase of the transcript per Million mapped reads (FPKM) plotted on a log10 transformed scale. High degree of correlation between the pairs indicates that similar transcript (gene) quantities are consistently detected in the samples (reproducibility of the method).

3.2.1.5. Transcriptome assembly/ mapping

Two different methods are used to produce a transcriptome from the raw sequence reads, i.e. 1. *de novo* assembly and 2. mapping to a reference genome. The first approach does not rely on the presence of the reference genome in order to reconstruct the nucleotide sequence. Due to the small size of the short reads, *de novo* assembly may be difficult. The second approach relies on the reference genome to map millions of reads. Recently specialized algorithms for transcriptome alignment have been developed, e.g. Bowtie for RNA-Seq short read alignment [65], TopHat for aligning reads to a reference genome to discover splice sites [115], Cufflinks to assemble the transcripts and compare/ merge them with others [116]. These tools can also be combined together to form a comprehensive system.

3.2.1.6. Sequencing data analysis

RNA-Seq data allows the characterization of a gene expression in cells or tissue via measurement of mRNA levels. For instance it can be determined how cells differ between a healthy or diseased state or at different developmental time points. Cuffdiff2 algorithm may be employed to estimate the abundance of transcripts (genes) and their differential expression in the samples. Gene expression levels are frequently normalized by the total number of mapped reads and expressed as Fragments Per Kilobase of the transcript per Million mapped reads (FPKM). Cuffdiff2 assumes that the expression of a transcript in each condition can be measured by counting the number of generated fragments. Thus, a change in the expression level of a transcript is measured by comparing its fragment count in each condition. The resulting p-values are corrected with Benjamini-Hochberg for multiple testing to get a meaningful FDR (False discovery rate, also called q-value). This value describes the probability that the observed expression difference represents a change when compared to the null hypothesis of no change. It is based on a statistical model evaluating measurement error, technical variability and cross-replicate biological variability [116].

Chapter 4

MATERIALS and METHODS

4.1. Animals

Lungs were harvested from wild type (WT) or transgenic mice. Mice were sacrificed by isoflurane overdose (cat. no.: 1001936060; Baxter, Deerfield, USA). All mouse lines were housed according to European Laboratory Animal Welfare Act and Regulations in animal housing facility in Max-Planck-Institute for Heart and Lung in Bad Nauheim. Mice were housed in a barrier facility with purified air and water, supplied with food and water ad libitum, and exposed to a 12:12-h light-dark cycle. All transgenic animals were maintained in C57BL/6 background. WT mice were purchased from Charles River. The file reference number from the ethics committee along with the permission for animal laboratory use, is as follow: B2/320 at the Regierungspräsidium Darmstadt.

Transgenic mouse lines:

a) B6.129S4-Pdgfra$^{tm11(EGFP)Sor}$/J; stock number: 007669; (PDGFRa-eGFP). *Pdgfrα* promoter drives constitutive expression of the H2B-eGFP fusion gene. Green fluorescence is detected in cell nucleus. The mouse line was donated by P. Soriano, Mount Sinai School of Medicine, USA. The mice were maintained as heterozygote. The line has been reported previously [42].

b) B6.FVB-Tg(Myh11-cre/ERT2)1Soff/J; stock number: 019079; (SmMHC/CreER). A CreERT2 sequence, a polyadenylation signal, and a frt-flanked beta-lactamase cassette were inserted into the initiation codon of *Myh11* (smooth muscle myosin, heavy polypeptide 11). The transgene is integrated on the Y chromosome. The Line was donated by Prof. Dr. S. Offermanns, Max-Planck-Institute for Heart and Lung, Germany. It has been described previously [124]. To activate CreERT2 recombinase, a dose of 0.5mg tamoxifen per neonate was injected intraperitoneally at 3 subsequent days before harvesting the lung. Tamoxifen stock solution of 20mg/ml: tamoxifen powder (cat. no.: T5648, Sigma-Aldrich, St. Louis, USA) dissolved in normal corn oil

by shaking for about 1h at room temperature. Tamoxifen was stored at -20°C protected from light.

c) STOCK Gt(ROSA)26Sor$^{tm4(ACTB-tdTomato,-EGFP)Luo}$/J; stock number: 007576; mTom/mGFP (mT/mG). The mice possess loxP sites on either side of a membrane-targeted tdTomato (mT) cassette and express a strong red fluorescence in all tissues and cell types. The line was bred to a mouse line expressing Cre recombinase. The offspring had the mT cassette deleted in the cre expressing tissue(s) after tamoxifen treatment. mT cassette deletion activated the expression of membrane-targeted EGFP (mG) cassette located downstream (Figure 10). Mice were purchased from the Jackson Laboratory and maintained as homozygote. The line has been reported previously [86].

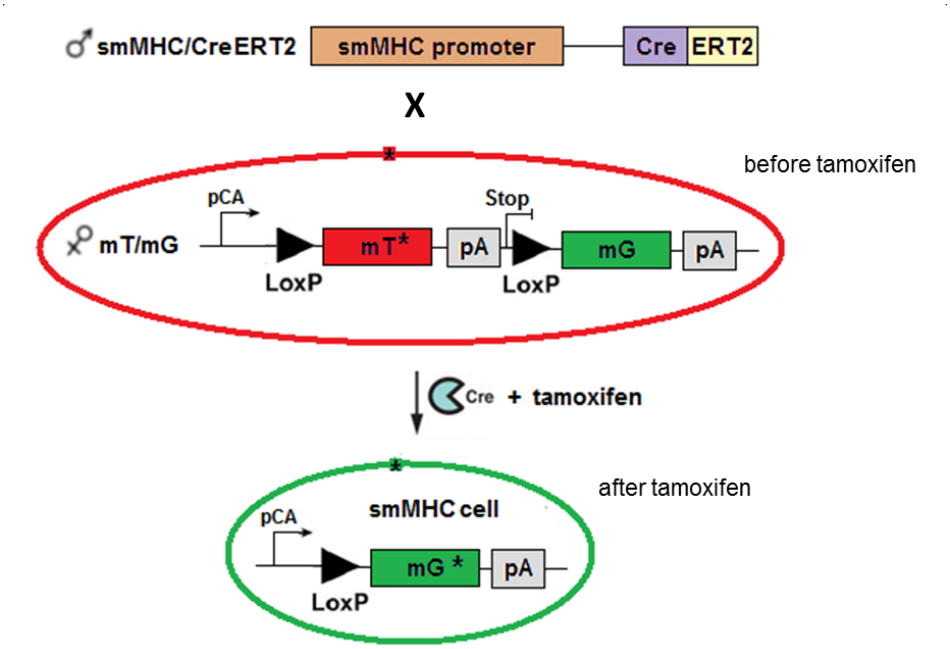


Figure 10 Schematic diagram of double transgenic SmMHC/CreERT2 and mT/mG mouse. *Cre-mediated recombination (tamoxifen inducible) causes mT deletion. As a result SmMHC-positive cells are characterized by membrane green fluorescence (mG; mGFP, *). * = active fluorescence.*

d) B6.Cg-Tg(Myh11-cre,-EGFP)2Mik/J; stock number: 007742; smMHC/Cre/eGFP. Smooth muscle myosin heavy chain (smMHC or Myh11) promoter directs bicistronic Cre and EGFP (enhanced green fluorescent) protein expression. Hemizygotes from founder line SMCG2 (SM2Cre/GFP) display intense EGFP fluorescence restricted to vascular and nonvascular smooth muscle with strong concordance between cre

expression and EGFP fluorescence. The mouse line was obtained from the Jackson Laboratory and maintained as heterozygote. The line has been reported previously [126].

4.2. Genotyping

DNA isolation from tail cuts

DNA was isolated from mouse tails obtained 1-2 weeks after birth. Tails were placed in 1.5ml tubes (cat. no.: 72.706; Sarstedt, Nümbrecht, Germany) and digested in 500µl lysis buffer (pH=8, 50mM Tris/HCL; 100mM EDTA; 100mM NaCl; 1% SDS) + 5µl proteinase K (stock solution 20mg/ml) (cat. no.: 7528.1; Roth, Karlsruhe, Germany). Samples were incubated at 55°C shaking overnight. They were centrifuged next day for 10min, at 13000rpm, 4°C (Eppendorf Centrifuge 5430R, Hamburg, Germany). Supernatants were collected into new 1.5ml tubes. 500µl of isopropanol (cat. no.: 6752.2; Roth, Karlsruhe, Germany) was added to supernatants and centrifuged for 10min at 13000rpm, 4°C. DNA pellets were washed with 500µl of 70% ethanol (cat. no.: 603-002-00-5; Merck, Darmstadt, Germany) and centrifuged for 10min at 13000rpm, 4°C. DNA pellets were dried and 100µl of T 1/10 E buffer (pH=8, 50mM Tris/HCl; 100mM EDTA) was added. Samples were incubated at 55°C in a shaking thermoblock overnight or 2h.

PCR and electrophoresis

2µl of DNA from each sample was used for semi-quantitative PCR followed by agarose gel electrophoresis. PCR mixture was prepared with 5µl of ImmoMix™ (cat. no.: BIO-25020; Bioline, Luckenwalde, Germany) master mix, 0.25µl appropriate primers (purchased from Eurofins) and supplemented by ddH_2O up to 10µl. Reaction mix was prepared in 0.2ml polypropylene 8-tube strips (cat. no.: 673271, 683271; GBO, Frickenhausen, Germany).

Following primers were used for genotyping of the specific mouse lines:

a) PDGFRa-eGFP (product size: mutant = 242bp, wild type = 451bp):

5’ - CCC TTG TGG TCA TGC CAA AC - 3’ wild type forward

5’ - GCT TTT GCC TCC ATT ACA CTG G - 3’ wild type reverse

5’ - ACG AAG TTA TTA GGT CCC TCG AC - 3’ mutant reverse

b) SmMHC/CreER (product size: transgene = 287bp, internal positive control = 324bp):
   5’ - TGACCCCATCTCTTCACTCC - 3’ SMWT1
   5’ - AACTCCACGACCACCTCATC - 3’ SMWT2
   5’ - AGTCCCTCACATCCTCAGGTT - 3’ phCREAS1
c) mTom/mGFP (product size: mutant = 250bp, wild type = 330bp):
   5’ - CTC TGC TGC CTC CTG GCT TCT - 3’ wild type forward
   5’ - CGA GGC GGA TCA CAA GCA ATA - 3’ wild type reverse
   5’ - TCA ATG GGC GGG GGT CGT T - 3’ mutant reverse
d) smMHC/Cre/eGFP (transgene = ~100bp, internal positive control = 324bp):
   5’ - GCG GTC TGG CAG TAA AAA CTA TC - 3’ transgene
   5’ - GTG AAA CAG CAT TGC TGT CAC TT - 3’ transgene
   5’ - CTA GGC CAC AGA ATT GAA AGA TCT - 3’ internal positive control forward
   5’ - GTA GGT GGA AAT TCT AGC ATC ATC C- 3’ internal positive control reverse

The PCR products were run on agarose gel (cat. no.: 2267.4; Roth, Karlsruhe, Germany) containing of 4µl ethidium bromide (10mg/ml; cat. no.: 2218.2; Roth, Karlsruhe, Germany) in 1xTAE buffer (40mM Tris acetate and 1mM EDTA) at 120V for 50-60min. The DNA products size separation was visualized under UV lamp. As a size marker 100bp ladder (cat. no.: M-214; Jena Bioscience, Jena, Germany) was used.

4.3. Primary cells isolation

Mouse primary cells were isolated from PBS-perfused and shredded lungs. Following lysis mixture was used: serum-free Dulbecco’s modified Eagle’s medium (DMEM) + GlutaMax (cat. no.: 31966-021; Life Technologies, Carlsbad, USA) containing 0.2% collagenase B (0.243U/mg; cat. no.: 11088815001; Roche, Penzberg, Germany), 10 U/cm^2 dispase (5000 caseinolytic units, cat. no.: 354235; BD, Franklin Lakes, USA) and 0.5µg/µl DNase (4065.6U/mg; cat. no.: 18535; Serva, Heidelberg, Germany). Lungs shredded on petri dishes were shaken for 30-40min at 37°C, 5% CO_2. Enzymatic activity of lysis reaction was stopped by adding DMEM/F12 medium (cat. no.: 21331-020; Life Technologies, Carlsbad, USA) containing 10% fetal calf serum (FCS; cat. no.: A15-101)

and 1% penicillin/streptomycin (cat. no.: P11-010) both purchased from PAA, Pasching, Austria. Undigested remains were eliminated by filtering the cell suspension through a cell strainer with a 100-µm nylon mesh (cat. no.: 352360; BD, Franklin Lakes, USA). The singe-cells suspension was centrifuged (Heraeus Multifuge 1S; Thermo Scientific, Rockford, USA) at 2000rpm, 10min, RT in a 50-ml polypropylene centrifuge tube (cat. no.: 227261; GBO, Frickenhausen, Germany). Red blood cells were lysed in lysis buffer [pH=7.4; 0.15M NH_4Cl (cat. no.: K298.1), 10mM $KHCO_3$ (cat. no.: P748.1), 0.1mM EDTA (cat. no.: 8040.3); all purchased from Roth, Karlsruhe, Germany], for 5min in RT. Pelleted (1500rpm, 5min, RT) cells were washed successively with Dulbecco's PBS (cat. no.: H15-002; PAA, Paschin, Austia), and centrifuged at 1500rpm, 5min. Cell pellets were frozen at -80°C or directly used for further experiments. Cell number was established microscopically (Zeiss Primo Vert; Oberkochen, Germany) using a Neubauer haemocytometer chamber (cat. no.: 0640010; Marienfeld, Lauda-Königshofen, Germany).

4.4. Fluorescent-activated cell sorting (FACS)

Cells sorting was performed using FACSAria™ III cell sorter with BD FACSDiva™ software v6.1.3 (BD Bioscience, San Jose, USA). Cell suspension was kept in 2.5ml sorting buffer [0.1% BSA (cat. no.: K45-001; PAA, Pasching, Austria), 2mM EDTA (cat. no.: 8040.3; Roth, Karlsruhe, Germany]. Samples in polystyrene tubes (cat. no.: 352052; BD Falcon, San Jose, USA) were transported on ice within 30min to the Institute for Clinical Immunology and Transfusion Medicine facility in Giessen. 100-µm nozzle and argon-ion laser (488-nm) were used to sort the cells. Cells were sorted into DMEM medium (cat. no.: 31966-021; Gibco, Grand Island, USA) within ca. 2h. Post-sorting test was run using sytox blue. The cells viability and purity was measured. Sorted cells were pelleted and total RNA was isolated.

4.5. Lung primary cells culture

Lung primary cells were cultured in DMEM/F12 medium (cat. no.: 21331-020; Gibco, Grand Island, USA) supplemented with 10% FCS (cat. no.: A15-101) and 1% penicilin/streptomycin cat. no.: P11-010) both purchase from PAA, Pasching, Austria. Cells were seeded on 100mm (3.5×10^6 cells), 60mm (1×10^6 cells) polystyrene culture dishes (cat. no.: 664160 and 628160; GBO, Frickenhausen, Germany) or chamber slides (1.3×10^5 cells) (cat. no.: 354108; BD Bioscience, San Jose, USA). Cells were incubated at 37°C, 5% CO_2 conditions for 2, 24, 48 or 72h.

4.6. Positive and negative cell selection using magnetic Dynabeads

Dynabeads sheep anti-rat IgG (cat. no.: 110.35; Invitrogen, Carlsbad, USA) were used according to the manufacturer's instruction. Rat anti-PDGFRα (cat. no.: ab90967; Abcam, Cambridge, USA) and rat anti-CD45 (cat. no.: 550539; BD Bioscience, San Jose, USA) antibodies were used to separate PDGFRα-positive cells or leucocytes, respectively. Leucocytes were removed from samples using negative selection. Beads-unbounded cells were transferred to new tubes containing beads/anti-PDGFRα antibody complexes in sorting buffer (PBS with 0.1% BSA, 2mM EDTA). Cells were incubated in 5-ml polystyrene tubes (cat. no.: 115 101; GBO, Frickenhausen, Germany) on ice by gently mixing time to time. After 30min of incubation cells were washed 3 times with sorting buffer. Separated cells were centrifuged for 5min at 10000rpm, 4°C. Cell pellets were directly frozen in RNA later at -80°C (cat. no.: AM7020; Ambion, Carlsbad, USA).

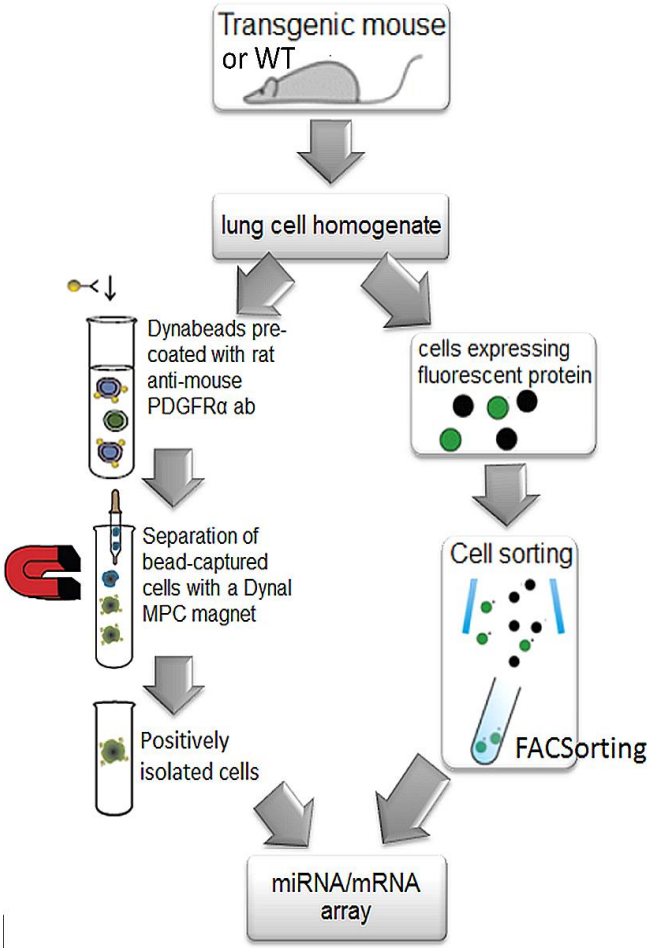


Figure 11 Principle of magnetic Dynabeads cell sorting. *Bead-sorted cells were used for total RNA purification. In comparison FACSorting method was shown. MPC= Magnetic Particle Concentrator.*

4.7. Immunofluorescence, confocal and fluorescence microscopy imaging

4.7.1. Tissue preparation

PBS-perfused lungs were inflated intratrachealy with 1:1 PBS (cat no.: H15-002; PAA, Paschin, Austia) and Tissue-Tek (cat. no.: 4583; Sakura, Torrance, USA). Removed lungs were embedded in Tissue-Tek and immediately frozen in 2-methylbutane (cat. no.: 3927.1; Roth, Karlsruhe, Germany) kept on dry ice. Lungs from smMHC/CreER-mT/mG

and smMHC/Cre/EGFP transgenic mice were inflated intratrachealy with 4% PFA and incubated overnight in 4% PFA at 4°C. Next day, the lungs were washed with PBS for 2h at 4°C, embedded in Tissue-Tek and frozen in 2-methylbutane on dry ice. 10µm lung cryosections were cut using Leica CM3050S cryostat/microtome and adhered on polysine adhesion slides (cat. no.: 10143265; Thermo Scientific, Waltham, USA). Tissue slides were kept at -20°C.

4.7.2. Immunohistochemistry and immunocytology

10µm lung tissue slides were fixed with 1:1 mixture of methanol (cat. no.: 4627.5) and aceton (cat. no.: 9372.2) (both from Roth, Karlsruhe, Germany) within 10min at RT. Slides were washed 2-3 times with PBS (137mM NaCl, 2mM KH2PO4, 10mM Na2HPO4 · 2H2O, pH=7.4). Nonspecific binding of antibodies was blocked with 1:1 histobuffer (3% BSA in 1x PBS with 0.2% Triton X-100 cat. no.: 3051.3; Roth, Karlsruhe, Germany) and goat serum (cat. no.: 16210-072; Life Technologies, Carlsbad, USA) within 1h at RT. All antibodies were diluted in histobuffer and incubated in black box filled with water at RT. Primary antibodies were added directly after blocking and incubated on the slides for 1h. Secondary antibodies were added and incubated for 45-60min. If the staining was double or triple the steps were repeated. Nuclei were stained for 20min at RT with DAPI (1:1000; cat. no.: D1306), To-Pro3 (1:2000; cat. no.: T3605) or Sytox blue (1:1000; cat. no.: S11348), all three obtained from Life Technologies, Carlsbad, USA. After each staining, slides were washed extensively with PBS. After last staining, tissue slides were incubated in 4% PFA (cat. no.: 441244; Sigma-Aldrich, St. Louis, USA) for 5min and washed with PBS. The same steps were applied for cells staining in chamber slides; all incubation times were shortened by half. Slides were mounted with mowiol (cat. no.: 475904; Merck, Darmstadt, Germany).

1:300 rabbit anti- ADRP polyclonal antibody (cat. no.: ab52356) and 1:300 rat anti-PDGFR alpha monoclonal antibody (cat. no.: ab 90967) from abcam (Cambridge, USA) were used. 1:500 mouse IgG2a (cat. no.: M5409), 1:500 mouse anti- actin α-smooth muscle-Cy3 (cat. no.: C6198) and 1:500 mouse anti- vimentin (cat. no.: V 6630) monoclonal antibodies purchased from Sigma-Aldrich, St. Louis, USA were used. 1:300 rat

IgG2a (cat. no.: 553927), 1:500 rat anti- CD45 (cat. no.: 550539) 1:500 and rat anti- CD31 (cat. no.: 550274) monoclonal antibodies from BD Biosciences, Franklin Lakes, USA were used. 1:500 Rabbit anti- cytokeratin polyclonal antibody (cat. no.: Z0622; Dako, Glostrup, Denmark), 1:500 rabbit anti- collagen I polyclonal antibody (cat. no.: T 40777R; Rockland, Gilbertsville, USA), 1:500 goat anti- GFP-FITC antibody (cat. no.: NB100-1771; Novus Biologicals, Littleton, USA), 1:300 rabbit IgG polyclonal antibody (cat. no.: 2729, Cell Signaling Technology, Danvers, USA), 1:700 mouse IgG2b-Cy3 (cat. no.: 0104-12; Southern Biotech, Birmingham, USA) were also used. Alexa Fluor (AF) 488 (cat. no.: A11008) or AF 555 (cat. no.: A21428) goat anti- rabbit IgG, AF 488 (cat. no.: A11006) or AF 555 (cat. no.: A21434) goat anti- rat IgG, AF 488 (cat. no.: A11001) goat anti- mouse IgG secondary antibodies were used and purchased from Invitrogen, Carlsbad, USA, all of them in 1:1000 dilution.

Zeiss LSM 710 confocal microscope (Carl Zeiss, Oberkochen, Germany) with ZEN 2011 Microscope Software or Leica DFC360 FX Fluorescent microscope (Leica Microsystem; Wetzlar, Germany) with Leica Application Suite (LAS) AF microscope software were used.

4.8. Total RNA isolation, real-time PCR

Total RNA was purified with GeneJET™ RNA Purification kit (cat. no.: K0731; Fermentas Thermo Fisher Scientific, St. Leon-Rot, Germany) according to the manufacturer's instructions. 100ng of total RNA was used for cDNA synthesis. cDNA was synthesized using High Capacity cDNA Reverse Transcription Kit (cat. no.: 4368814; Applied Biosystems, Foster City, USA) according to the manufacturer's instructions. 0.2-ml polypropylene 8-tube strips (cat. no.: 673271, 683271; GBO, Frickenhausen, Germany) were used. Eppendorf Mastercycler ep Gradient S thermal cycler (Eppendorf, Hamburg, Germany) was used.

cDNA synthesis protocol (reaction volume 20µl):

1: 25.0°C for 10min

2: 37.0°C for 120min

3: 85.0°C for 5min

4: 4.0°C hold

Real-time PCR was performed using a SensiMix SYBR No-ROX kit (cat. no.: QT650-05; Bioline, Luckenwalde, Germany) according to the manufacturer's instructions. 2µl of cDNA was used. The primers used to amplify mouse *Pdgfrα* transcript were: forward 5′ GGAGCCTGAGCTTTGAGCGACG 3′ and reverse 5′ GAAAGACCTGGTGGGAGG TCCC 3′. The primers for mouse *Adrp* were: forward 5′ CGACGACACC GATGAGTCCCAC 3′ and reverse 5′ TCAGGTTGCGGGCGATAGCC 3′. The primers for mouse *Myh11* (*Smmhc)* were: forward 5′ CAGCGCATCAACGCCAACCG 3′ and reverse 5′ AGCCTCGTTTCCTCTCCTGGTGG 3′. Mouse porphobilinogen deaminase (*Pbgd*) transcript was used as an internal control. The primers for *Pbgd* were: forward 5′ ATGTCCGGTAACGGCGGC 3′ and reverse 5′ GGTACAAGGCTTTCAGCATCGC 3′. All the primer pairs were separated by at least one intron to exclude genomic DNA contamination and checked for primer efficiency. Bio-Rad C1000 thermal cycler with CFX96 system (Hercules, USA) was used.

Real-time PCR protocol (reaction volume 15µl):

1: 95.0°C for 6:00 min

2: 95.0°C for 0:10 sec

3: 59.0°C for 0:10 sec

4: 72.0°C for 0:15 sec

Plate Read

5: GOTO 2; 44 more times

6: 95.0°C for 0:10 sec

7: Melt Curve 65°C to 95°C: Increment 0.5°C for 0:03 sec

Plate Read

4.9. Protein isolation, western blot

Cells were lysed in 1x RIPA lysis buffer (cat. no.: 89901) supplemented with 1:100 protease and phosphotase inhibitor cocktail (cat. no.: 1861284), and 1:100 EDTA solution (cat. no.: 1861283; all purchase from Thermo Scientific, Rockford, USA). Samples were centrifuged (Eppendorf Centrifuge 5430R, Hamburg, Germany) for 30min at 13000rpm,

4°C. Supernatants were collected into new 2-ml tubes. Proteins concentration was measured using a *Dc* protein assay kit (cat. no.: 500-0113; Bio-Rad, Hercules, CA, USA) according to the manufacturer's instructions and read using Nano Quant Infinite M200 (Tecan, Männedorf, Switzerland). 30µg of total protein was solubilized in loading buffer (375mM Tris-HCl pH=6.8, 10% SDS, 50% glycerol, 12.5% β-mercaptoethanol, 0.02% bromophenol blue) for 5min at 95°C. Proteins were resolved by 7.5% or 10% SDS-PAGE gel (375mM Tris/HCl, 7.5% or 10% acrylamide, 0.1% SDS, 0.05% APS, 0.1% TEMED) at 100V. 1x running buffer (25mM Tris, 192mM glycine, 0.1% SDS) was used. Gel was transferred to nitrocellulose membrane (cat. no.: S80209; Pall Corporation, Dreieich, Germany) at 100V for 70min in cooled transfer buffer (25mM Tris, 192mM glycine, 10% methanol). The membrane was blocked within 1h at RT with 5% skim milk (cat. no.: M7409; Sigma-Aldrich, St. Louis, USA) prepared in 1x TBST (50mM Tris, 150mM NaCl, 0.1% tween 20). After blocking, the membranes were blotted with specific antibodies diluted in 5% skim milk: anti-PDGFRα (1:1000), anti-α-SMA (1:1000), anti-ADRP (1:600) and anti-β-actin (1:5000). Antibodies were incubated for 1h or overnight at RT or 4ºC. Secondary antibodies (1:3000) were incubated for 45-60min at RT. After antibodies incubation, membranes were washed 3x for 10min with 1x TBST. If necessary a stripping buffer (cat. no.: 21062; Thermo Scientific, Rockford, USA) was used for 15min at RT. Protein expression was detected by chemiluminescence using Supersignal West Femto Max Sensitivity substrate (cat. no.: 34096; Thermo Scientific, Rocford, USA) and Lumi-Imager Fujifilm LAS-4000 (Tokyo, Japan). Rabbit anti-PDGFRα C-terminal polyclonal antibody (cat. no.: SAB4502141), mouse anti-actin α-smooth muscle monoclonal antibody (cat. no.: A5228) and mouse anti-β-actin monoclonal antibody (cat. no.: A5316) were purchased from Sigma-Aldrich (St. Louis, USA). Rabbit anti-ADRP polyclonal antibody (cat. no.: ab52356) was purchased from abcam (Cambridge, USA). Anti-rabbit IgG- peroxidase (POD) conjugated secondary antibody was purchase from Thermo Scientific (cat. no.: 31460; Rockford, USA) and anti-mouse IgG- POD conjugated secondary antibody from Sigma-Aldrich (cat. no.: A9917; St. Louis, USA).

4.10. Total RNA integrity and quality

RNA integrity was estimated using Agilent 2100 bioanalyzer (serial no.: DE24802342, Agilent, Santa Clara, USA) with 2100 Expert software. Agilent RNA 6000 Pico (cat. no.: 5067-1513) or Nano (cat. no.: 5067-1511) kits were used according to the manufacturer's instruction. RNA integrity number (RIN) was considered as a guidelines of RNA integrity. The RIN was estimated by the entire electrophoretic trace of the RNA sample (gel images next to the histograms). Smaller 28S/18S ratio values were associated with lower RIN, indicating slight RNA degradation.

4.11. MicroRNA TaqMan low density array (TLDA)

Total RNA was purified with RNeasy Micro Kit or RNeasy Mini Kit (cat. no.: 74004 or 74104 Qiagen; Venlo, Netherlands) according to the manufacturer's instruction. >350ng (suspended in 3µl) of input total RNA was used for cDNA synthesis. If necessary the RNA samples were pulled and precipitated with 1:10 3M sodium acetate, pH=5.5 (cat. no.: AM9740; Life Technologies, Carlsbad, USA) and 2.5 volume of 100% ethanol within 20min on ice. Samples were centrifuged for 10min at 13000rpm, 4°C. RNA pellets were washed with 75% ethanol and centrifuged for 5min at 13000rpm, 4°C. Pellets were dried at RT and dissolved in 4µl of nucleases free water to reach the desirable amount of concentrated RNA. RNA concentration was measured using Qubit 2.0 Fluorometer (cat. no.: Q32866) and Qubit RNA Assay Kit (cat. no.: Q32852) obtained from Life Technologies, Carlsbad, USA. cDNA was synthesized using TaqMan® MicroRNA Reverse Transcription Kit (cat. no.: PN 4366596) and Megaplex™ RT Primers Rodent Pool Set v3.0, pools A and B (cat. no.: 4444746) from Applied Biosystems (Foster City, USA) according to the manufacturer's instructions. Megaplex™ RT Primers were designed to convert mature miRNAs into cDNA. Eppendorf Mastercycler ep Gradient S thermal cycler (Eppendorf, Hamburg, Germany) was used.

cDNA synthesis thermal-cycling conditions (reaction volume 7,5µl):

1: 16°C for 2min

2: 42°C for 1min

3: 50°C for 1sec

4: GO TO step 1, 40 cycles

5: 85°C for 5min

6: 4°C hold

The cDNA was stored at -20°C.

TaqMan microRNA real time PCR arrays were performed using 384-well plate TaqMan® Array Rodent MicroRNA A+B Cards Set v3.0 (cat. no.: 4444909) and TaqMan® Universal PCR Master Mix II, No AmpErase® UNG, 2x (cat. no.: PN 4324018) from Applied Biosystems (Foster City, USA) according to the manufacturer's instructions. Comprehensive coverage of Sanger miRBase v15 was enabled across a two-card set of TaqMan® Array MicroRNA Cards (Cards A and B) for a total of 641 (mouse) and 373 (rat) unique assays. Card A array v2.0 and card B array v3.0, each contains 384 TaqMan® MicroRNA assays, enabling accurate quantitation of 335 (card A) and 306 (card B) unique microRNAs for mouse (remaining miRNAs consist of human and rat specific but identified in mouse as well). Together 750 specific miRNAs excluding controls were analyzed. On each card three mouse endogenous control assays (4x mammalian U6 snRNA, mouse specific snoRNA135 and snoRNA202, rat specific Y1 and U87) to aid in data normalization, and one assay not related to mammalian as a negative control (Arabidopsis thaliana ath-miR159a) were included. Card A contains better characterized miRNAs, while Card B contains many of the more recently discovered miRNAs along with the miR* sequences.

The TaqMan low density arrays were performed in collaboration with Prof. Dr. Bernd Schmeck and Dr. Wilhelm Bertrams at Department of Molecular Pneumology of Philipps University in Marburg. The Applied Biosystems® ViiA™ 7 Real-Time PCR System with TaqMan® Array Block (Low Density Array Thermal Cycling Block) and SDS software v2.1 (cat. no.: 4453537) was used. Each card was run in triplicates.

Each miRNA Ct value was normalized to 4x mammalian U6 snRNA, and calculated as follow: Average(not normed value U6snRNA_1, not normed value U6snRNA_2, not normed value U6snRNA_3, not normed value U6snRNA_4) - not normed value of miRNA of interest = ΔCt. All significant candidates from card A&B were showed as a median from

3 replicates. Heat map visualization and hierachical clustering was performed using Perseus software version 1.4.0.2. One-way ANOVA test was used; differences were considered statistically significant at $p<0.05$. Data analyses were performed in collaboration with biostatistician Dr. Mario Looso from Max Planck Institute in Bad Nauheim.

4.12. Next-Generation Sequencing

Next-Generation Sequencing was performed using Ion Proton™ System (includes Ion Proton™ Sequencer and Ion Proton™ Torrent Server) (cat. no.: 4476610; Life Technologies, Carlsbad, USA). Total RNA was purified with RNeasy Micro Kit or RNeasy Mini Kit (cat. no.: 74004 or 74104 Qiagen; Venlo, Netherlands) according to the manufacturer's instruction. Ribosomal RNA (cytoplasmic and mitochondrial) was depleted using Low Input RiboMinus™ Eukaryote System v2 (cat. no.: A15027; Life Technologies, Carlsbad, USA) according to the manufacturer's instruction. RiboMinus™-enriched RNA samples were fragmented by RNaseIII to the desired size of 100-200bp fragments and transcribed into cDNA using Ion Total RNA-Seq Kit v2 (cat. no.: 4475936; Life Technologies, Carlsbad, USA) according to the manufacturer's instruction. Each cDNA library was created from 2 combined RNA samples (according to Figure 12). The Ion PI™ Template OT2 200 Kit v2 (cat. no.: 4485146; Life Technologies, Carlsbad, USA) was used for template preparation. Templated Ion PI™ Ion Sphere™ particles were prepared for sequencing runs by use of Ion PI™ Sequencing 200 Kit v2 (cat. no.: 4485149; Life Technologies, Carlsbad, USA) and further deposited in Ion PI™V2 chips (cat. no.: 4482321; Life Technologies, Carlsbad, USA). The sequencing quality control and replicates correlation were performed.

Differential expression of isoforms of the respective samples was analyzed using the TopHat-Cuffdiff pipeline:

1. Creation of a Bowtie 2 genome index for reference mouse assembly version mm9 (NCBIM37)
2. Alignments of RNA-Seq reads to the genome to identify exon-exon splice junctions using TopHat including annotation from the NCBI Annotation Release 103

3. Reference-guided assembly of RNA-Seq reads into transcripts and report of abundance estimates (RPKM/FPKM) using Cufflinks

4. Comparison of assembled transcripts with a reference annotation, or across experiments, using Cuffcompare

5. Identification of statistically significant differences in transcript expression using Cuffdiff2.

All tools were run with default parameters. The final lists were filtered for a minimum FDR (false discovery rate) < 0.05 and log2FC (fold change) <> -1/1.

FPKM = Fragments Per Kilobase of transcript per Million mapped reads,

FPKMx = FPKM of the gene in sample x; FPKMy = FPKM of the gene in sample y;

log2(FPKMy/FPKMx) = the (base 2) log of the fold change y/x

q-value = the FDR-adjusted p-value of the test statistic after Benjamini-Hochberg correction for multiple-testing.

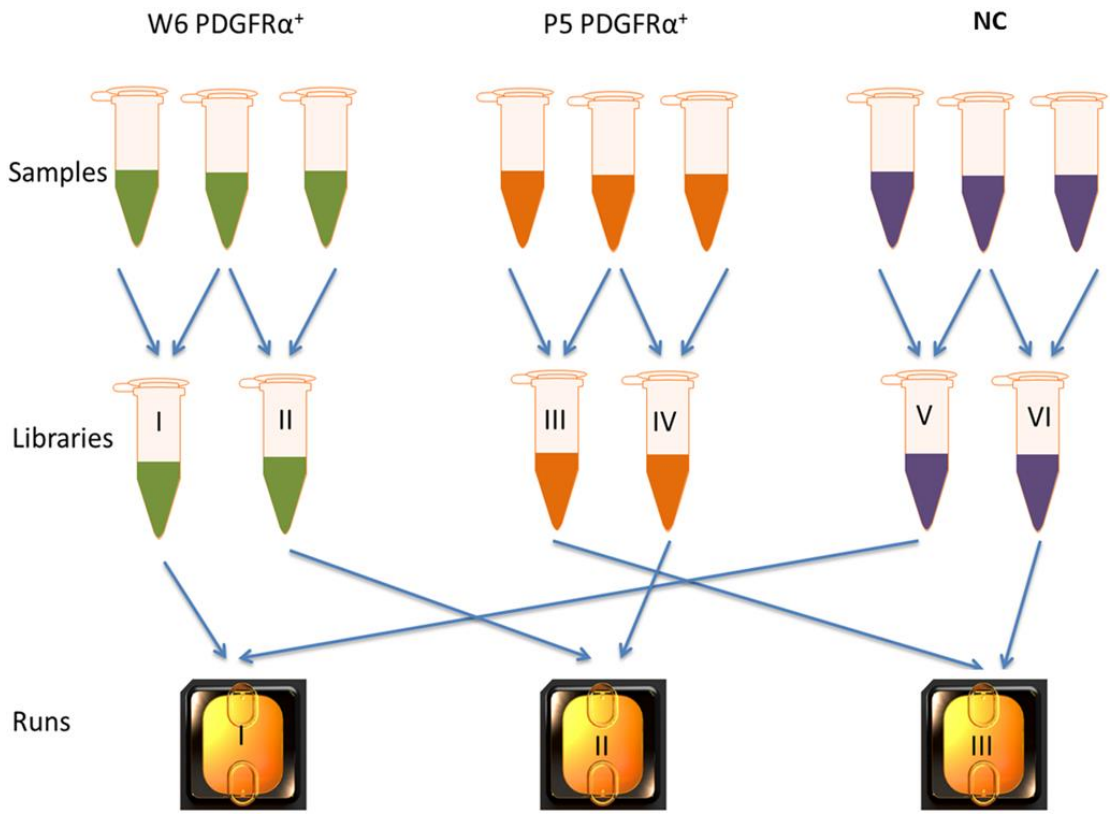


Figure 12 Scheme of PDGFRα-positive cells transcriptome sequencing experiment setup. *Three RNA samples were used for each time point (W6, P5) and negative control (cells from whole lung homogenate without CD45- and PDGFRα-positive cells). One cDNA library was prepared from*

two samples. Two independent cDNA libraries from each time points or NC were loaded on the chips. NC = negative control.

Next generation sequencing was performed by Dr. Stefan Günther within the ECCPS Next-generation sequencing and bioinformatics platform at the Max Planck Institute in Bad Nauheim. Analysis of the data was done in collaboration with biostatisticians Dr. Carsten Künne and Dr. Mario Looso from Max Planck Institute in Bad Nauheim/ ECCPS Next generation sequencing and bioinformatics platform.

4.13. Statistical analysis

The results were shown as mean ± SEM for each condition unless otherwise stated in description. Significance of the parametric differences among more than 2 groups was evaluated using one-way ANOVA and between 2 groups with Student's t-test using GraphPad Prism 5 (GraphPad Software, Inc., USA) or Microsoft Excel 2010 (Microsoft, Redmond, USA) respectively. All experiments were repeated at least 3 times, and representative experiments were shown. Differences were considered statistically significant at $p<0.05$. Pictures were assembled using ImageJ 1.46r. Representative pictures of at least 3-times repeated experiments are shown.

4.14. Computational analysis

For the computational data analyses the following software (algorithms) was used:

a) TargetScan Release 5.2: prediction of miRNA targets database developed at Whitehead Institute for Biomedical Research.
b) miRTarBase Release 4.4: experimentally validated microRNA-target interactions database; Department of Biological Science and Technology, Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu, Taiwan [50].
c) Ingenuity Pathway Analysis (IPA): allows exploring, interpreting, and analyzing complex biological systems. MiRNA functions and molecular interactions analysis

were based on experimentally observed datasets (based on number of reviewed publications). P-values were generated for each function using Fisher's exact test. Citing: Ingenuity® Systems, www.ingenuity.com.

d) GOrilla (updated March 8th 2013): a tool for identifying and visualizing enriched gene ontology (GO) terms in ranked lists of genes. Components of the GOrilla system were designed as part of the European Union FP6 funded Multi Knowledge Project [36].

e) Molecular Signatures Database (MSigDB) v4.0, updated May 31, 2013: collection of annotated gene sets for use with Gene Set Enrichment Analysis (GSEA) software. GSEA is a computational analysis of the overlaps between experimental gene set and gene sets in MSigDB. The gene sets are curated from the following online databases:

Name	**URL/Reference**
KEGG (Kyoto Encyclopedia of Genes and Genomes)	http://www.genome.jp/kegg
Pathway Interaction Database (PID)	http://pid.nci.nih.gov
Reactome	http://www.reactome.org
BioCarta	http://www.biocarta.com
Signal Transduction KE (ST)	http://stke.sciencemag.org
Signaling Gateway	http://www.signaling-gateway.org

GSEA is a method that determines whether an a priori defined set of genes shows statistically significant differences between two biological states concerning their differential expression (phenotypes, time points, etc.). It considers experiments with genome-wide expression profiles from samples belonging to two classes. Enrichment score (ES) reflects the degree to which a gene set is overrepresented at the extremes (top or bottom) of the entire ranked list. The score is calculated by walking down the list, increasing a running-sum statistic when a gene is encountered in one gene set (correlated with phenotype) and decreasing it when genes are not encountered in this gene set. The ES must be adjusted to account for differences in the gene set sizes. Gene sets that are in the middle of ranked list i.e. are not specific for one phenotype are accounted in an additional step of normalization (ES is normalized by accounting the size of each gene set). Therefore, a normalized enrichment score (NES) allows to compare analysis results across gene sets.

Leading edge gene subsets show gene members that contribute most to the enrichment score (after or before the peak of enrichment score). The nominal p-values need to be corrected to adjust for multiple hypothesis testing. False positives results are controlled by calculating the false discovery rate (FDR) corresponding to each NES. The FDR stands for the estimated probability that a set with a given NES represents a false positive finding. It is computed by comparing the tails of the observed and null distributions for the NES. A significantly enriched gene set is considered if its NES has an FDR q-value below 0.25 or the family wise error rate (FREW; probability of making one or more false discoveries) is <0.05. The statistical stringency of FREW is higher than FDR (q-value). Gene Set Enrichment Analysis (GSEA) was developed at the Broad Institute of MIT and Harvard [92].

Chapter 5

RESULTS

5.1. Characterization of fibroblasts subsets in lung tissue

Based on specific phenotypic markers of previously identified lung fibroblasts subsets, their temporal and spatial expression pattern was evaluated within wild type or transgenic mice lung tissue across different postnatal time points (P3, P5, P7, P10, P14) and in adult (W8) mice (Figure 13). ADRP, the phenotypic marker of lipofibroblast was linearly down-regulated during postnatal lung development and slightly detectable in adult tissue (Figure 13). ADRP-positive cells were co-stained with vimentin (fibroblast marker), CD31 (endothelial cell marker) and CD45 (leukocyte marker). ADRP protein was depicted as numerous small dots within whole lung tissue besides the bronchial epithelium and the endothelium of big vessels (Figure 14A). Wild type mouse PDGFRα-positive cells were present within the alveolar interstitial tissue, around the bronchi and the adventitial space of the vessels (Figure 13). Transgenic PDGFRα-GFP-positive cells were co-stained with ADRP, αSMA (myofibroblast and smooth muscle cells marker), CD31, cytokeratin (CK; epithelium marker), and CD45 (Figure 15). PDGFRα-GFP-positive cells moderately co-expressed αSMA at P3 and highly at P14 but not in adult lung tissue. Moreover, PDGFRα-GFP-positive cells surrounding the bronchi expressed GFP on a lower level (dim) than in the alveolar interstitial walls (bright). αSMA expression abundantly increased after P5 and peaked between P7-P10 and then again decreased at P14. In the adult lung tissue, αSMA expression level was almost undetectable in the interstitium but still highly expressed around the vessels and bronchi (Figure 13). Additionally, smMHC/CreER-mT/mG and smMHC/Cre/eGFP transgenic mice lung tissues were used to identify myofibroblasts localization (Figure 14C). αSMA and SMMHC (mGFP- or eGFP-tagged) proteins were localized at the tips of the newly forming secondary septa and around the alveolar rings (Figure 14B). Isotype controls for primary and secondary antibodies were used for each staining (Appendix, Figure A2).

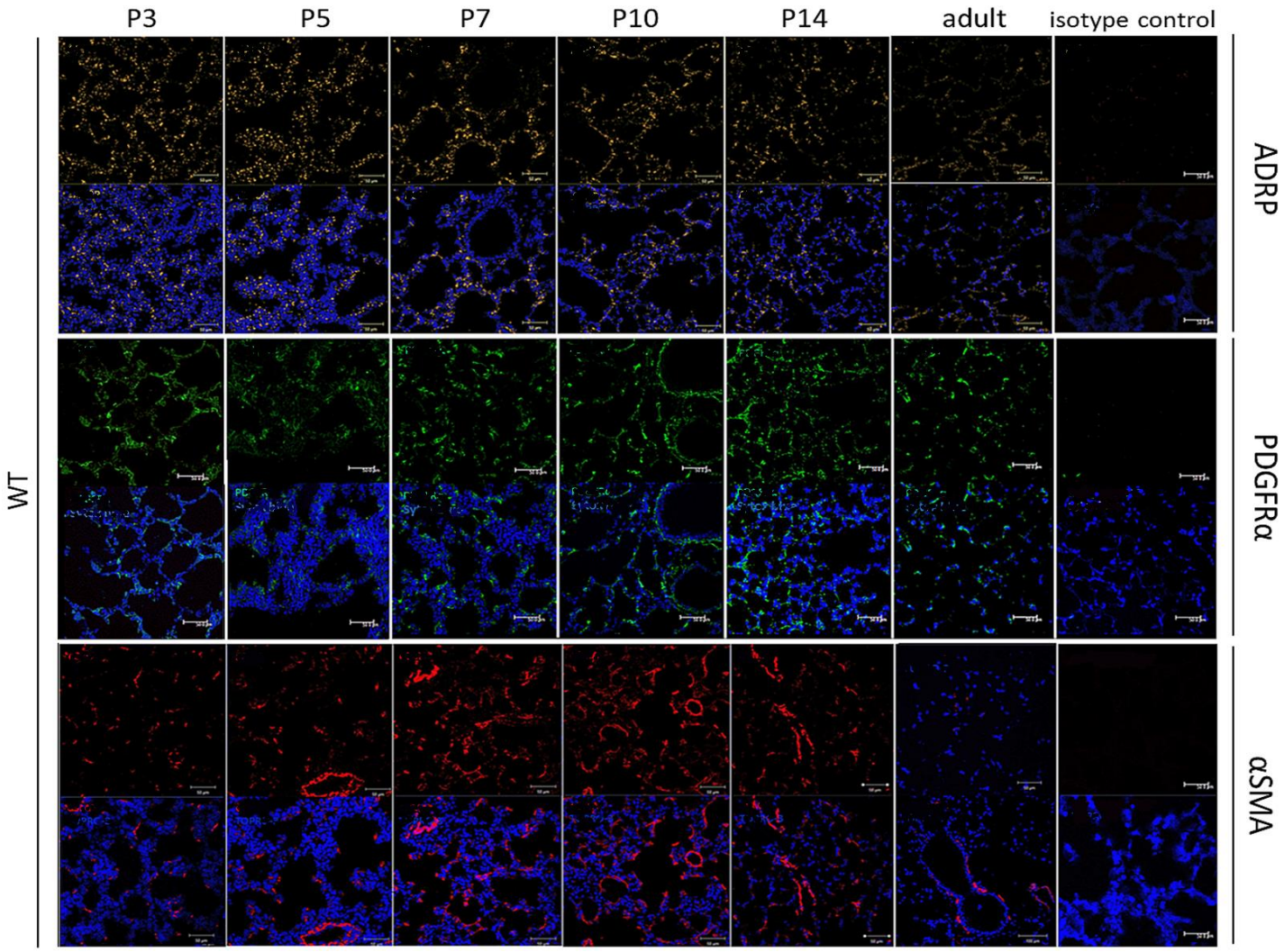


Figure 13 ***Temporal and spatial expression patter of lung fibroblasts subsets during postnatal developmental time points and in adult mice.*** *Wild type mouse lung tissue was fluorescently stained with antibodies against ADRP (yellow), PDGFRα (green) and αSMA (red) at postnatal day 3, 5, 7, 10, 14 and week 8. Isotype control was used to adjust fluorescent intensity for each marker. Anti -ADRP, - PDGFRα and - αSMA/Cy3 primary antibodies were used. Alexa fluor 488 secondary antibody was used. Nuclei were stained with To-Pro3 or Sytox blue. Scale bar = 50µm.*

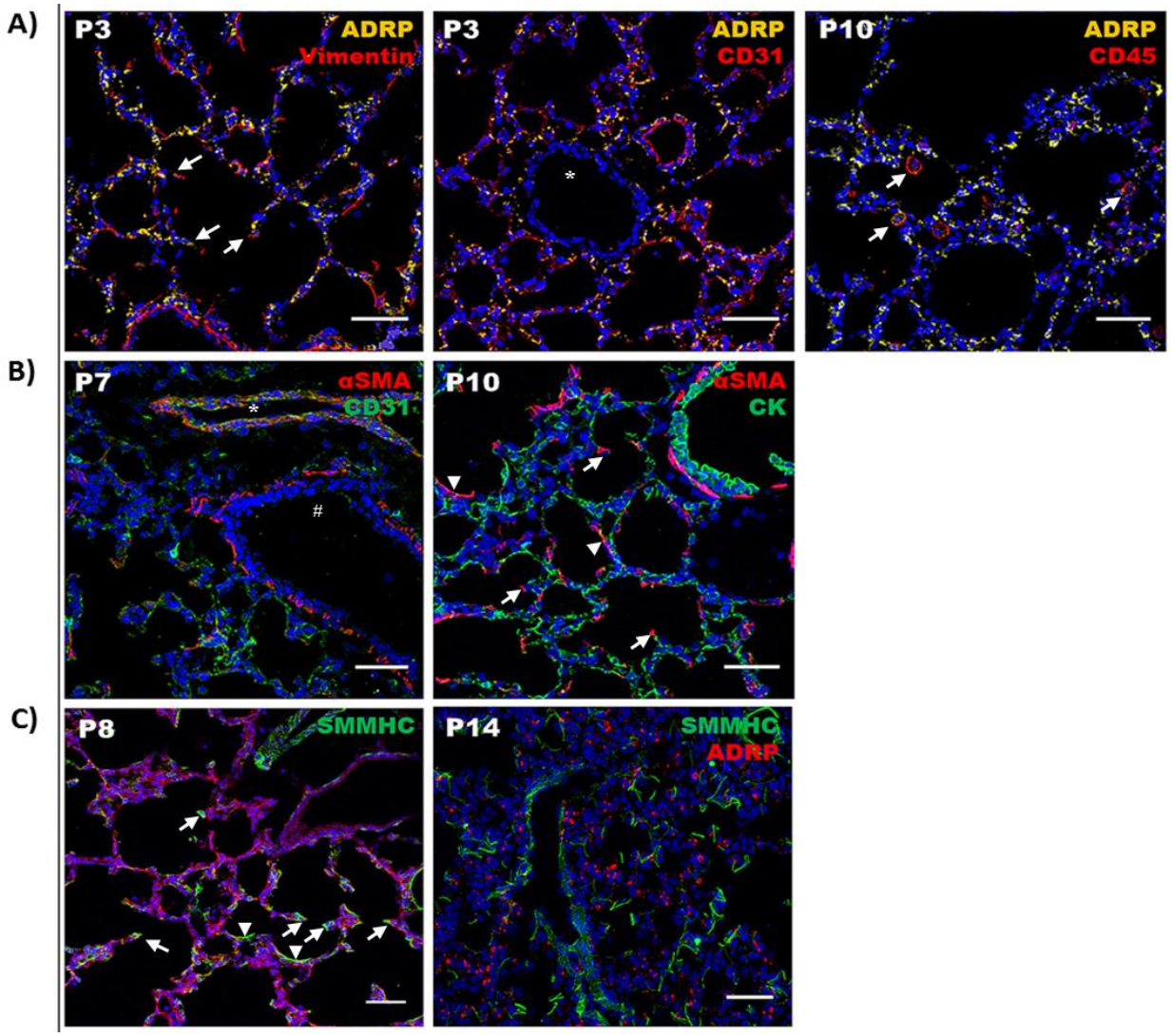


Figure 14 Localization of lipofibroblast and myofibroblast markers in postnatal mouse lung tissue (WT). *Anti -ADRP -vimentin, -CD31 and -CD45, -αSMA/Cy3, -CK primary antibodies were used. Alexa fluor 555 or 488 secondary antibodies were used. Nuclei were stained with To-Pro3 (blue). Scale bar = 50µm. Isotype controls are shown in Appendix, Figure A2.* ***A)*** *ADRP (yellow) is localized within whole lung tissue as small dots surrounding lipid droplets; it is not expressed by bronchial epithelium (*) but positively co-expressed by some CD45- (red) and vimentin- (red) positive cells (arrows).* ***B)*** *αSMA (red) is localized around vessels (*) and bronchi/bronchioles (#) as well as in the interstitium. In the interstitium it is localized at the tips of newly formed septa (arrows) and around the alveolar rings (arrow head)* ***C)*** *Tissue from double transgenic smMHC/CreER-mT/mG (P8) mouse lung is shown; SMMHC-positive cells (green) turned from red (mT) into green (mGFP) after tamoxifen-induced CreER-loxP recombination; SMMHC-negative*

cells remained red. SmMHC/Cre/eGFP (P14) transgenic mouse lung tissue was co-stained with ADRP (red). SmMHC localization is similar to αSMA. Anti -GFP/FITC antibody was used.

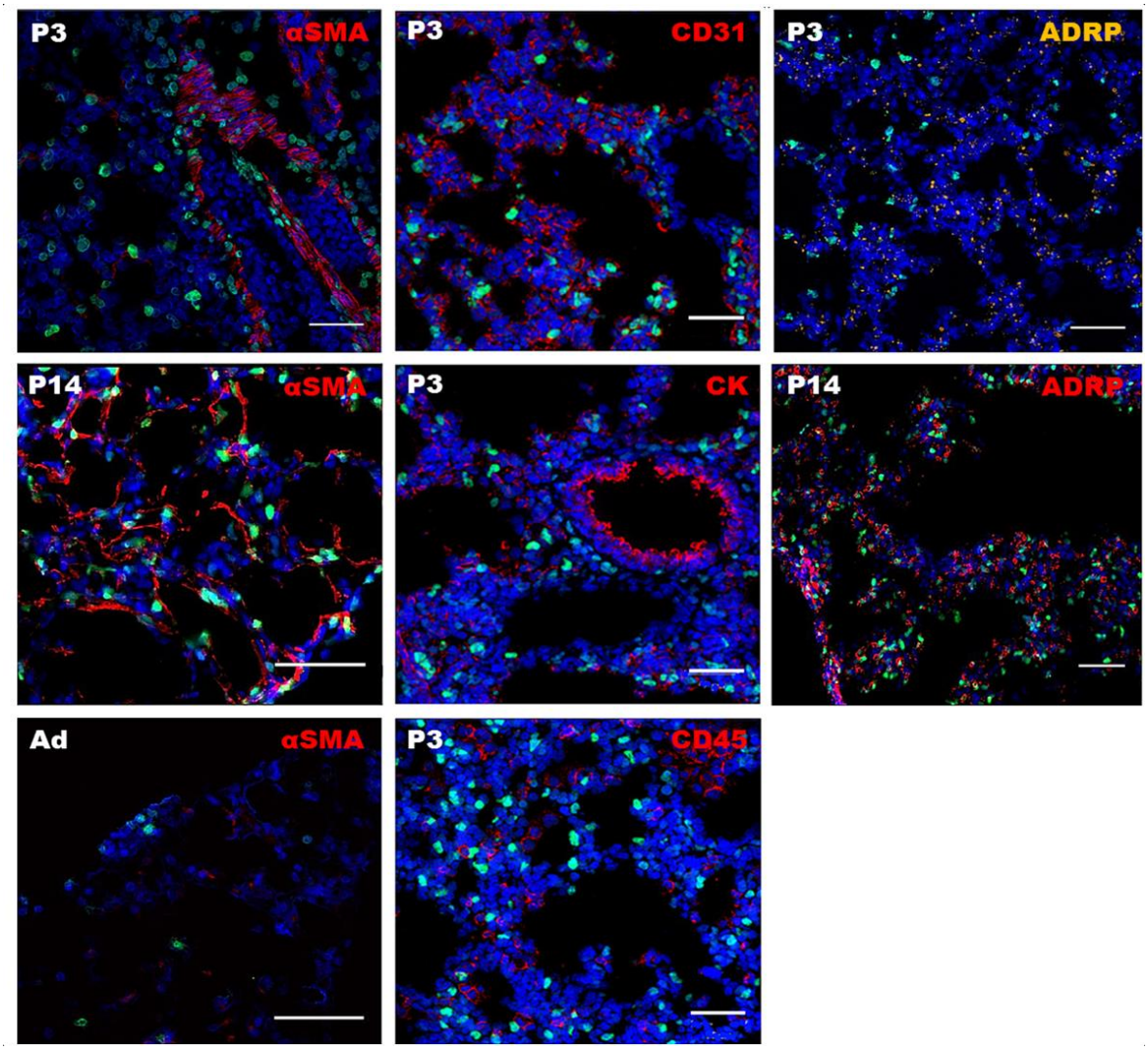


Figure 15 Localization of PDGFRα-positive cells in lung tissue. *PDGFRα-GFP transgenic mouse were used. PDGFRα-positive cells (green) are present in interstitium as well as parabronchial and adventitial space. Anti-GFP/FITC antibody was used to increase GFP signal expressed in the cell nuclei. Anti -αSMA/Cy3 (red), -ADRP (yellow or red), -CD31 (red), -CK (red) and -CD45 (red) primary antibodies and Alexa Fluor 555 secondary antibody were used. Nuclei were stained with To-Pro3 (blue). Scale bar = 50μm. Ad = adult. Isotype controls are shown in Appendix, Figure A2.*

Fibroblasts subsets phenotypic markers expression intensities were measured on gene and protein levels in whole lung homogenate. The mean of protein/β-actin densitometric ratio is shown on bar chart (Figure 16A). Expression patterns of ADRP,

αSMA (αSMactin) and PDGFRα proteins were similar to those obtained by tissue immunostaining (Figure 13). All three proteins were down-regulated in the adult mouse lung (week 8). ADRP protein expression level peaked at P5, αSMA at P10 and PDGFRα was highest at P2 then down-regulated at P5 and again slightly up-regulated. Replicates of protein blots from different littermates (2-5 animals) are also presented (Figure 16B). αSMA expression level obtained from each replicate was highly variable at P5, P10 and P14.

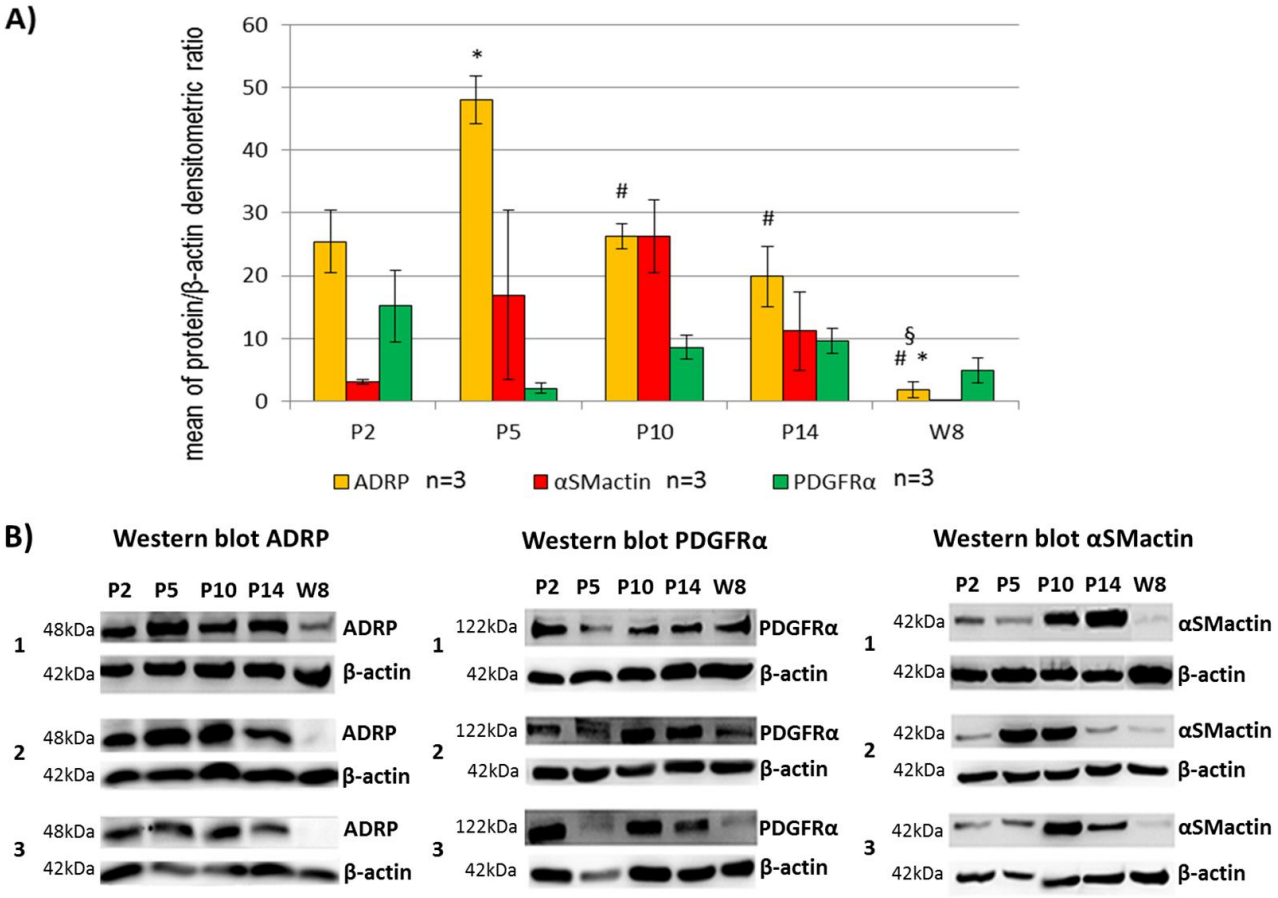


Figure 16 Protein expression levels in whole lung homogenate during different postnatal lung developmental time points. A) *Each bar shows mean ± SEM of protein/β-actin densitometric ratio from three (n=3) replicates (independent groups of animals, n= 2-5). One-way ANOVA (Bonferroni's post-test) was applied; p<0.05; statistically significant were considered: ADRP for: *P2 vs. P5, *P2 vs. W8, #P5 vs. P10, #P5 vs. P14, #P5 vs. W8, §P10 vs. W8.* ***B)*** *Replicated protein blots from different littermates are shown. Molecular mass of each protein is indicated as kDa. β-actin as loading control was used. P= postnatal day, W= week after birth.*

Expression level of *Adrp*, *Smmhc (approved gene symbol: MYH11)* and *Pdgfrα* genes was evaluated using real time PCR and normalized to *Pbgd* housekeeping gene (Figure 17). *Adrp* was regulated on a much higher level than *Smmhc* and *Pdgfrα*. *Pdgfrα* was significantly up-regulated at P14 and in the adult tissue (week 8) comparing to P2. *Adrp* was significantly up-regulated at P14 and W8 vs. P2, P5 and P10. Similar to the αSMA protein expression level, *Smmhc* peaked at P10 and its expression level showed high variability between the replicates.

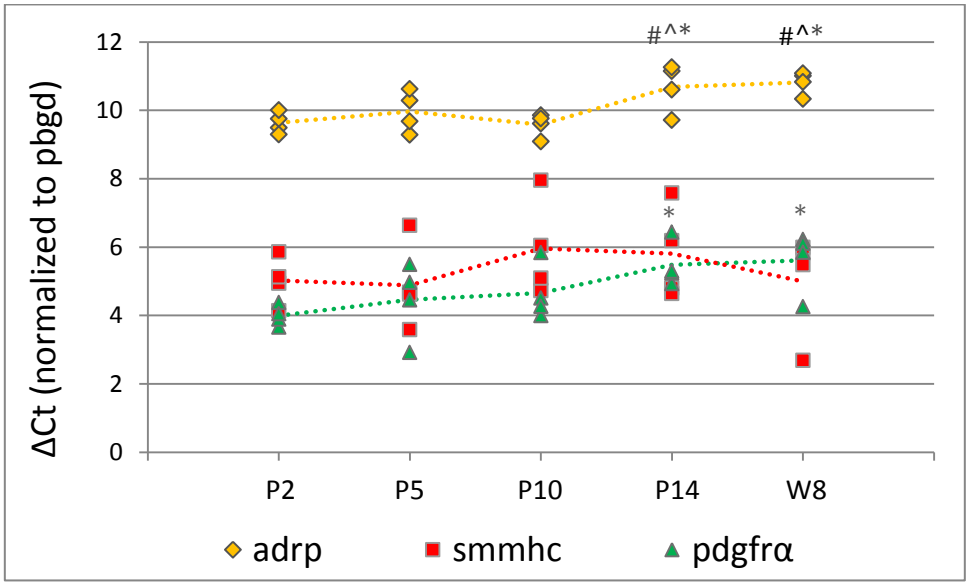


Figure 17 Gene expression levels in whole lung homogenate during different postnatal lung developmental time points. *Scatter plot chart shows expression level of Adrp, Smmhc and Pdgfrα normalized to Pbgd housekeeping gene (ΔCt = housekeeping gene – target gene). Four (N=4) independent groups of animals (n= 2-5 littermate) were used for each time point. One-way ANOVA (Newman-Keuls test) was applied; p<0.05; statistically significant were considered: pdgfrα: *P2 vs. P14, *P2 vs. W8; adrp: *P2 vs. P14, *P2 vs. W8, #P5 vs. P14, #P5 vs. W8, ^P10 vs. P14, ^P10 vs. W8. Each single symbol (square, triangle or rhombus) indicate ΔCt value from independent replicate.*

5.2. Characterization of myofibroblasts and lipofibroblasts in cell culture

Primary cells from lung homogenate were cultured in standard conditions and the fibroblasts phenotype was investigated after 1 or 2 days (Figure 18). Cells from postnatal day 6 (P6) and adult (W8) mouse lungs were co-stained with anti -αSM actin and anti -ADRP antibodies. After 1 day in cell culture, the cells from P6 lung were differentiated into αSM actin positive (αSM actin^{+}) myofibroblasts that co-expressed ADRP, and cells that expressed only ADRP. Cells isolated from the adult mouse lung differentiated into myofibroblast much slower than those from P6 and did not co-express ADRP after 1 day in cell culture. However the ADRP-single positive cells were present in the culture obtained from adult mouse. After 2 days of culture, almost all fibroblasts from the P6 lung were differentiated into double positive αSM actin^{+} ADRP^{+} cells. Myofibroblasts from adult lung also co-expressed ADRP at day 2. Isotype controls for primary and secondary antibodies were used (Appendix, Figure A1).

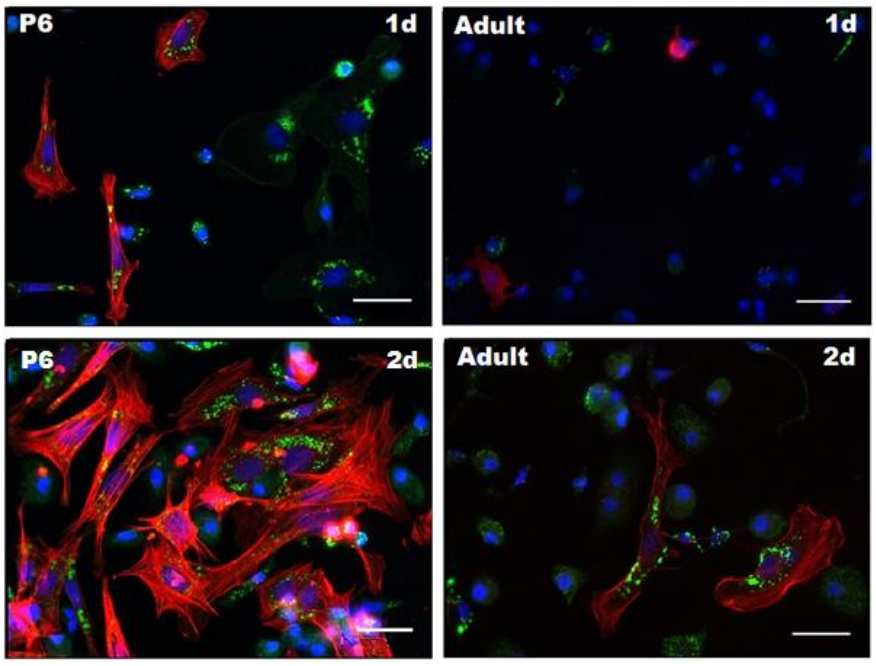


Figure 18 ADRP and αSM actin expression pattern in lung primary fibroblasts from WT mouse lungs after 1 and 2 days (1d and 2d) in standard cell culture. *Cells were isolated from postnatal day 6 (P6) or adult animals. Cells obtained from postnatal day 6 lung possess stronger*

differentiation capacities than cells from adult lung. Pictures are shown in merge, αSM actin in red, ADRP in green, and nuclei in blue (DAPI). Anti –αSM actin/Cy3, -ADRP primary antibodies and AF 488 secondary antibody were used. Observation was repeated more than 3 times at the same conditions. Scale bar = 30 µm. Isotype controls for primary and secondary antibodies are shown in Appendix, Figure A1.

5.3. Characterization of PDGFRα-positive fibroblasts in cell culture

To analyze the phenotype of PDGFRα-positive cells, PDGFRα-GFP transgenic mice were used. Lung primary cells from postnatal day 5 (P5) were cultured 3 days in standard conditions. Cells were stained with anti -αSM actin, -collagen I (Col I), -CD45, -ADRP, and -cytokeratin (CK) antibodies. PDGFRα-positive cells (GFP- tagged nuclei) co-expressed αSM actin, collagen I and ADRP but not CD45 and cytokeratin (Figure 19).

PDGFRα-positive cells were analyzed during their differentiation process into myofibroblasts in cell culture conditions. Observed changes of the PDGFRα expression level during the cell differentiation were based on GFP expression intensity level (Figure 20). The GFP intensity level was in average of 2.4-fold change lower in differentiated cells (cultured longer, 72 hours) than in undifferentiated cells (cultured 2 hours or 24hours) (Figure 20B). The GFP intensity was measured from 3 different wells of a chamber slide for each time point as an average intensity of the pixels in the marked area (green nuclei) using ImageJ. After 2 or 24 hours GFP protein expression intensity (bright) was more heterogeneous than after 72 hours with more homogenous expression level (dim). GFP intensity decreased accordingly with αSM actin cytoskeleton formation and cell size growth (Figure 20A). ADRP was expressed in both PDGFRα-GFP$^{bright+}$ and PDGFRα-GFP^{dim+} cells (Figure 20).

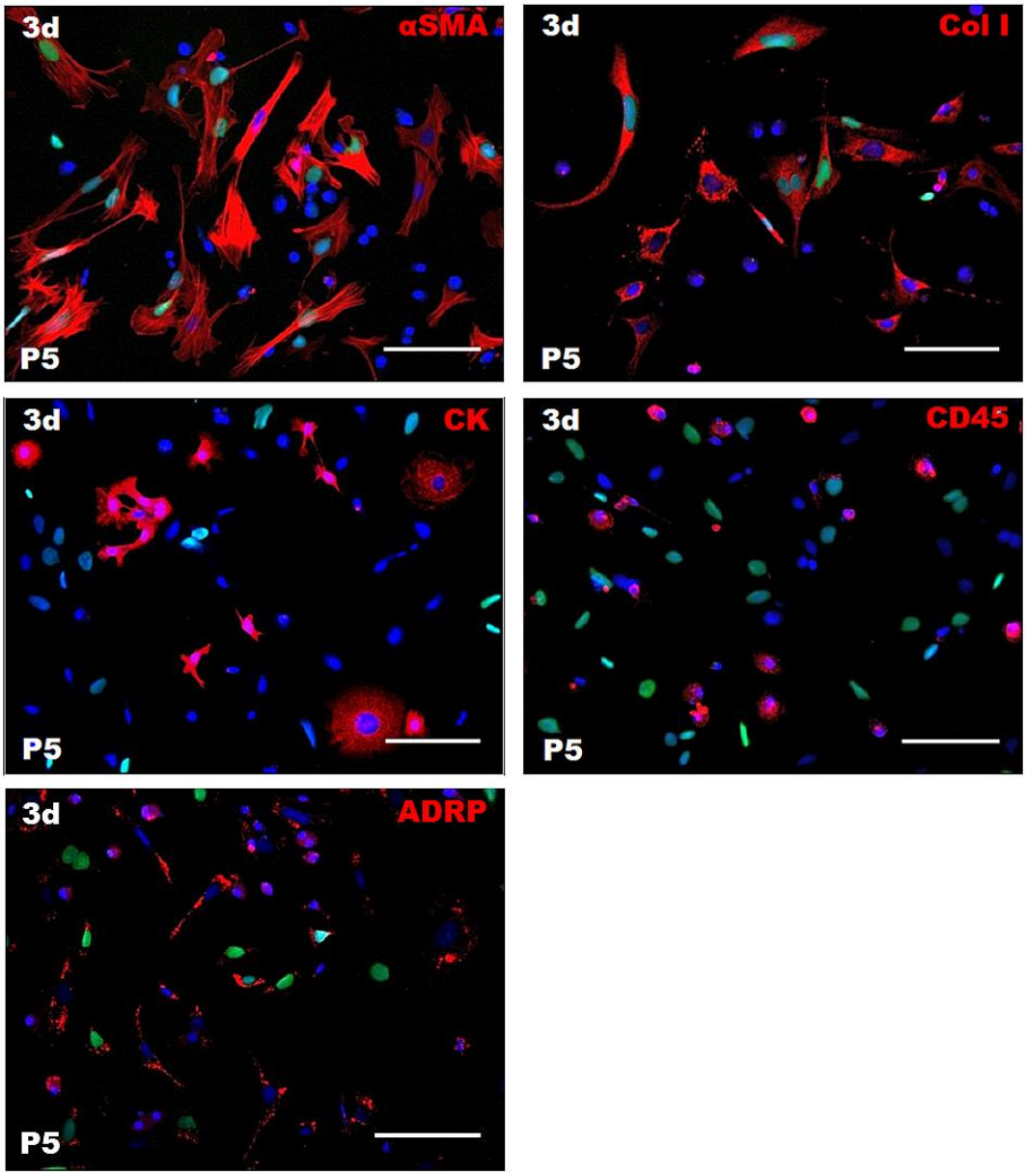


Figure 19 Phenotype of PDGFRα-positive cells. *Cells were obtained from a PDGFRα-GFP transgenic mouse at postnatal day 5 (P5). Cells were cultured 3 days (3d) in standard conditions. PDGFRα-positive cells are shown as GFP-positive (green). GFP-positive cells co-expressed αSMA, Col I and ADRP but not CD45 and CK. Anti -GFP/FITC, -αSM actin (αSMA)/Cy3, -collagen I (Col I), -cytokeratin (CK), -ADRP and -CD45 primary antibodies were used. Alexa Fluor 555 secondary antibody was used. Cell nuclei were stained with DAPI. Scale bar = 100µm. Isotype controls for primary and secondary antibodies are shown in Appendix, Figure A1.*

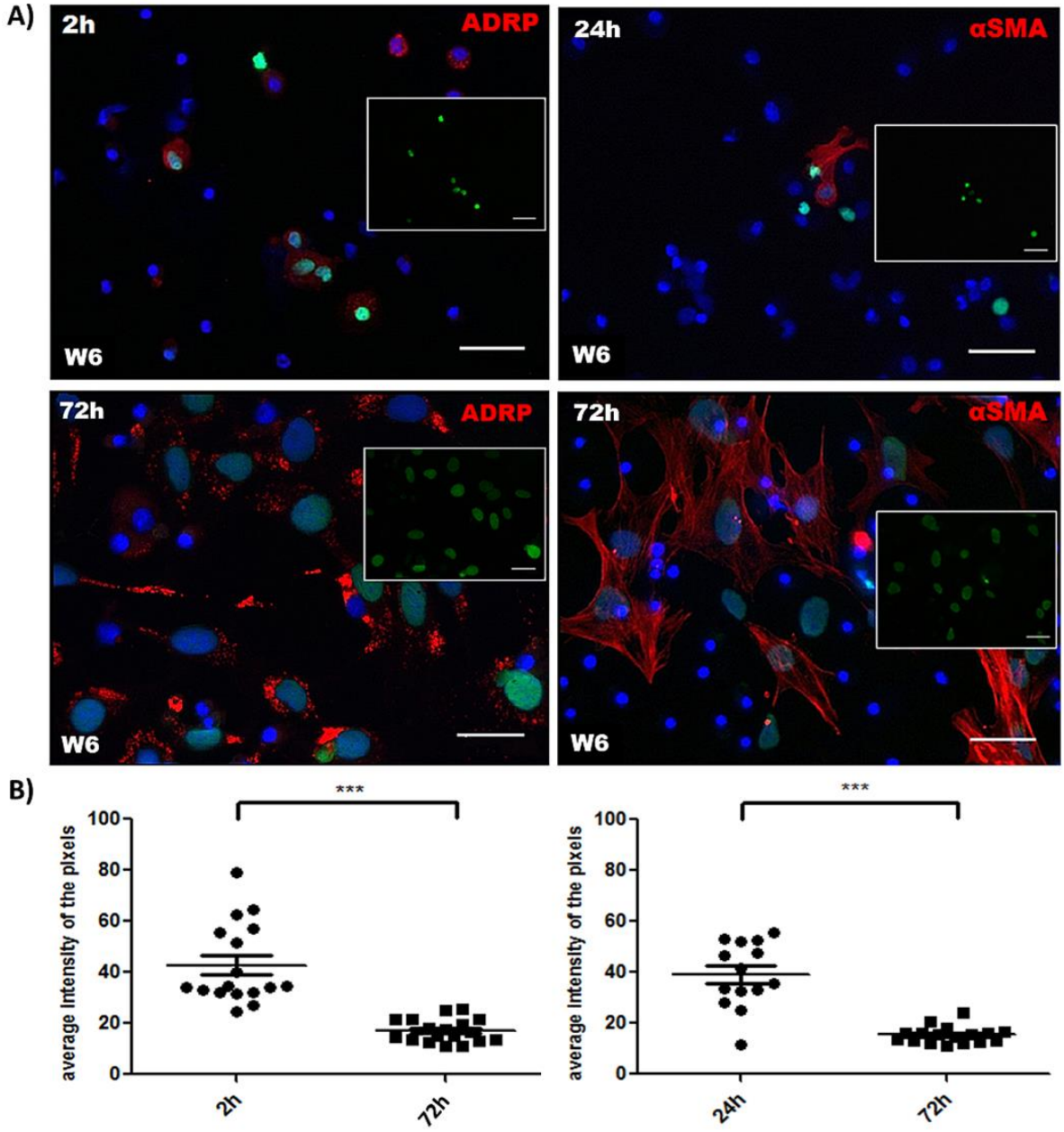


Figure 20 Changes of PDGFRα expression level intensity during PDGFRα-GFP-positive cells differentiation. A) *PDGFRα expression level was evaluated based on GFP intensity. Cells from week (W6) mice were cultured 2 hours (2h), 24 hours (24h) and 72 hours (72h). Anti -GFP/FITC, -ADRP, -αSM actin/Cy3 primary antibodies, and AF 555 secondary antibody were used. Insets show* GFP^+ *nuclei. Nuclei were stained with DAPI. Scale bar = 30µm. Isotype controls for primary and secondary antibodies are shown in Appendix, Figure A1.* ***B)*** *Average intensity of the pixels in the marked area (GFP-positive nuclei showed in inset) was measured using ImageJ. GFP intensity at different time points in cell culture was measured from 3 independent pictures taken with the same*

*microscope settings. Scatter plot graphs with mean +/- SEM are shown. Student's t-test, $p<0.05$ was applied. *** $p\leq0.001$.*

Cells isolated from P5 wild type mouse lungs were cultured 2 days and co-stained with anti -αSM actin and anti -PDGFRα antibodies (Figure 21). Fully differentiated αSM actin^{+} myofibroblasts were PDGFRα-negative. Undifferentiated PDGFRα$^{bright+}$ fibroblasts with long thin cytoplasmic extensions (filopodia) were starting to co-express αSM actin with very high intensity and forming α-actin cytoskeleton.

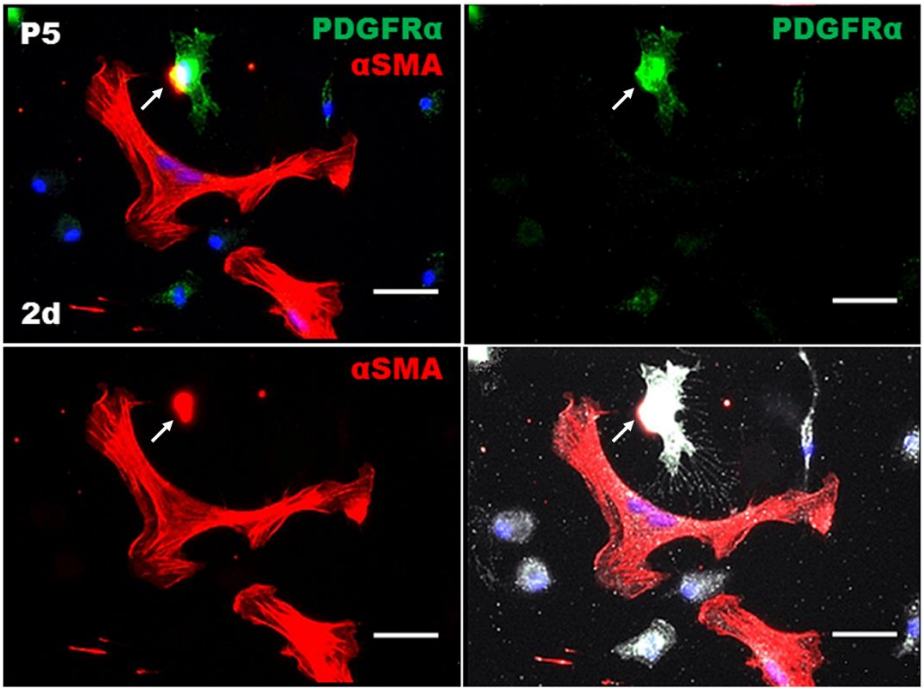
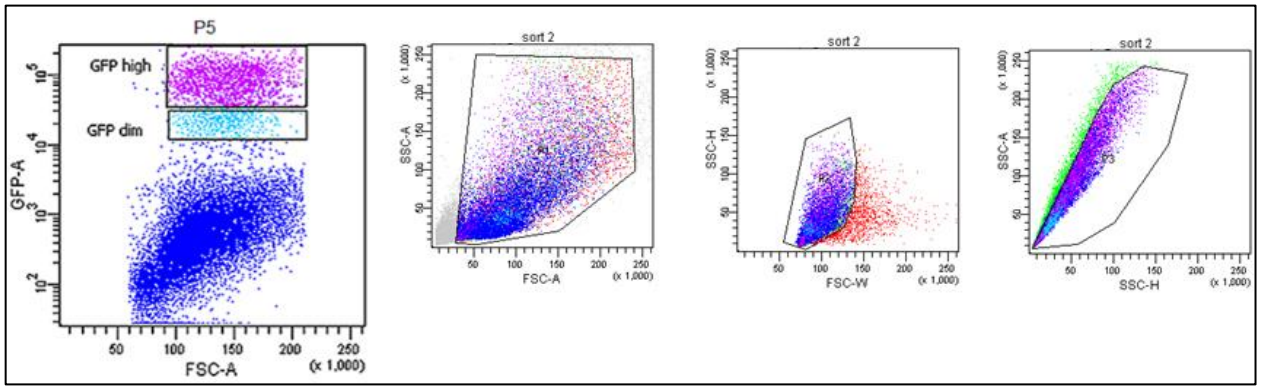


Figure 21 Differentiation of PDGFRα-positive fibroblast into myofibroblast. *Cells were isolated from WT P5 mouse lung homogenate. After 2 days in culture, cells were co-stained with anti- αSM actin/Cy3 and anti- PDGFRα primary antibodies. Undifferentiated PDGFRα$^{bright+}$ cells co-expressed αSM-actin with high expression intensity (arrow). Fully differentiated myofibroblast were αSM actin-positive but not PDGFRα-positive. AF 488 secondary antibody was used to detect anti-PDGFRα primary antibody. Nuclei were stained with DAPI. Experiment was repeated at least 3 times at the same conditions. Scale bar = 30µm. Isotype controls for primary and secondary antibodies are shown in Appendix, Figure A1.*

5.4. Fluorescence-activated cell sorting (FACS) of PDGFRα-positive cells

Cells isolated from PDGFRα-GFP transgenic mice were subjected to FACSorting (Figure 22). GFP-positive cells from postnatal day 5 mice constituted ~9% of the total cell number and consisted of 2 subpopulations: PDGFRα-GFP^{dim} (0.7% of total cells; Figure 22, upper panel) and PDGFRα-GFP$^{bright\ (high)}$ (8.2% of total cells; Figure 22, upper panel). GFP-positive cells obtained from mice 6 weeks after birth consisted of 1 population with very high GFP expression level and constituted ~3.2% of the total cell number (Figure 22, lower panel). Repeated post-sorting analysis of GFP-positive cells revealed a purity of ~96-98% and a viability of ~55-80% (Figure 22, middle panel). Though PDGFRα-GFP^{+} sorted cells showed a high purity they did not show a sufficient viability. Thus the RNA quality was not sufficient to continue with the further experiments of miRNA and mRNA profiling. Hence, a different cell isolation technique was chosen (Section 5.5.).



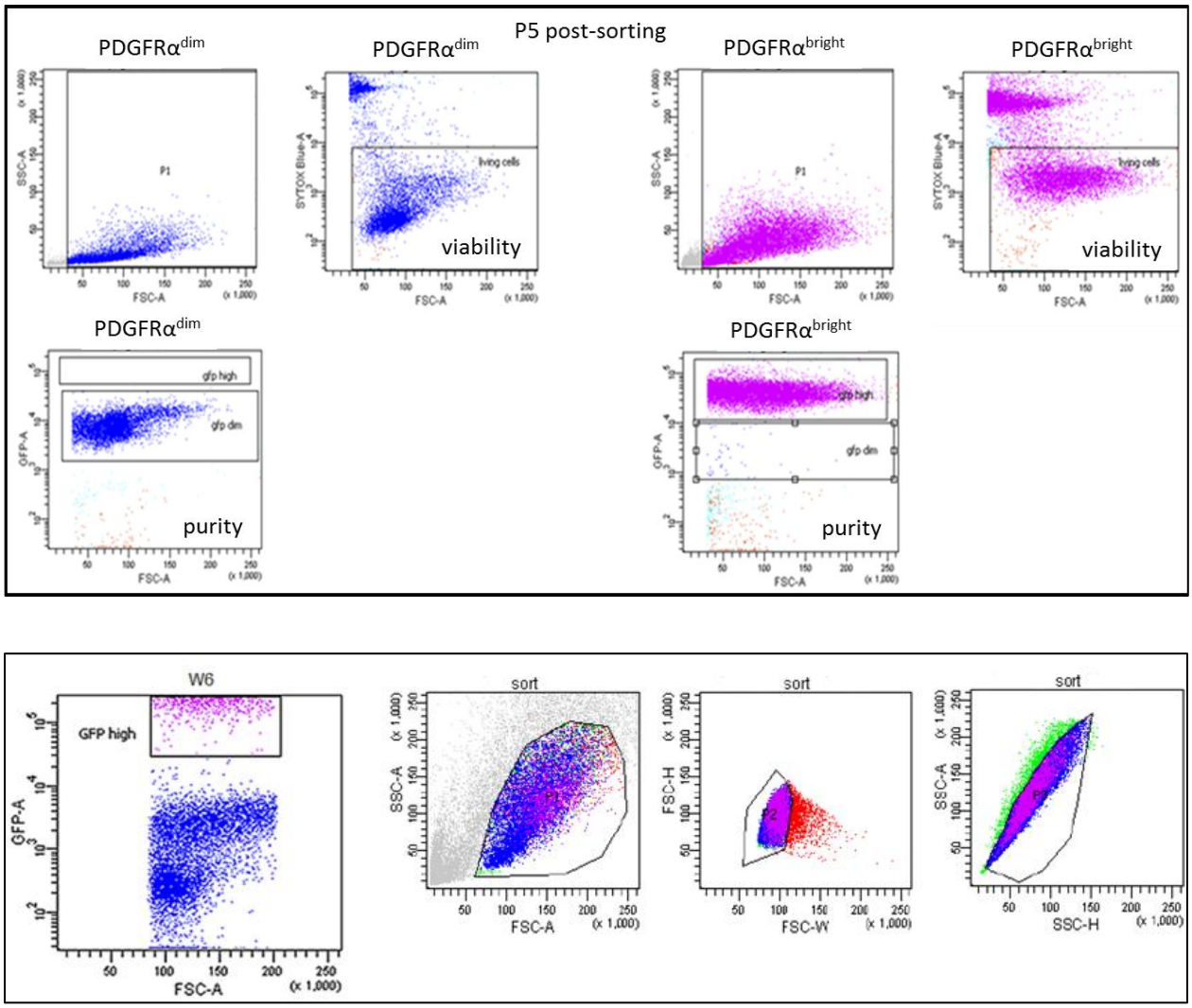


Figure 22 FACSorting of PDGFRα-GFP-positive cells from postnatal day 5 (P5) and week 6 (W6) lungs. *Cells were gated in the FSC-A vs. SSC-A channels to eliminate autofluorescent debris. The FSC-H vs. FSC-W and the SSC-A vs. SSC-H were used to depict doublet/aggregate discrimination. Upper panel shows cells sorted from P5 lungs. GFP-positive cells were gated as "dim" or "bright (high)" subpopulations. Middle panel shows post sorting of cells from P5 lungs. Viability was measured by counting the incorporation of sytox blue into dead cells, middle panel upper part. Cell purity is shown in the middle panel lower part. Lower panel shows cells sorted from W6 lungs where GFP-positive cells were gated as "bright (high)". BD FACSDiva™ software v6.1.3 was used.*

5.5. PDGFRα-positive cells separation by the use of magnetic Dynabeads

The isolation of PDGFRα-GFP-positive cells using the magnetic Dynabeads technique was chosen as alternative method for the previous described FACSorting of PDGFRα-positive cells (Section 5.4). To verify the beads positive sorting and purity, PDGFRα-GFP-positive cells were used. Sorted cells were cultured in standard conditions and visualized as GFP-positive cells surrounded by magnetic beads (Figure 23). Magnetic beads sorted cells were counted using Neubauer chamber. The number of GFP-positive cells was calculated as percentage of the total cell number sorted by use of magnetic beads. The obtained cell subset purity was ~ 85-90%.

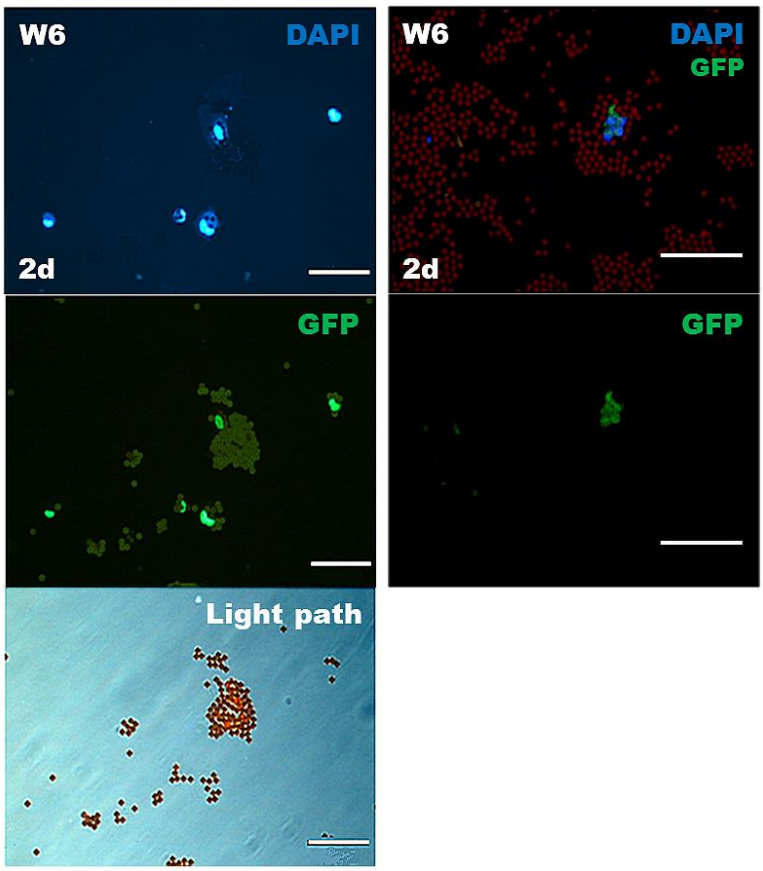


Figure 23 PDGFRα-GFP-positive cells separation based on magnetic bead sorting. *Sheep anti-rat magnetic beads were coupled with a rat anti -PDGFRα antibody. Positively sorted cells were culture 2 days in standard conditions; GFP-positive cells are surrounded by magnetic beads. Anti-GFP/FITC antibody was used. Nuclei were stained with DAPI. Scale bare = 50µm.*

5.6. Integrity and quality of PDGFRα-positive cells total RNA

The integrity of total RNA was evaluated using a bioanalyzer instrument. The RNA integrity number (RIN) ranged approximately from 7-8, whereas the ratio of 28S/18S rRNA subunit varied approximately from 1.3-2.5 (Figure 24**,** representative graphs). The RNA quality was considered as good enough for further miRNome analysis.

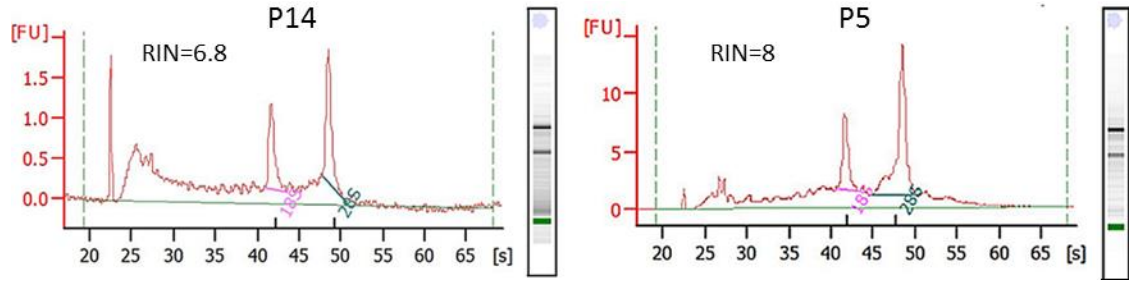


Figure 24 Total RNA integrity of sorted PDGFRα-positive cells. *The peaks of 28S and 18S rRNA subunits are marked on the histograms and visualized on gel images as darker bands (see also Figure 30). Y-axis shows fluorescence units (FU), x-axis shows time in seconds [s]. Marker is indicated as a green band on the gel image. Eukaryote total RNA Nano Kit was used. RIN = RNA Integrity Number.*

5.7. MicroRNA expression profiles of pulmonary PDGFRα-positive cells across different time points of lung development

PDGFRα-positive cells miRNA expression profiles across four different time points of lung development (E16.5, P5, P14 and W6) were generated by TaqMan low density array. Two array cards were used; card A and card B (see 4.11.). From 750 miRNAs only 98 were significantly regulated across four time points (one-way ANOVA test; $p<0.05$). The miRNA expression levels are presented on a heat map (Figure 25). A hierarchical clustering algorithm was used to visualize similarities of the expression patterns between these 98 regulated miRNAs. 2 major clusters were extracted. The upper cluster "1" with highly expressed miRNAs (brighter yellow) and the lower cluster "2" with general lower expression level across time points (darker blue and darker yellow). K-means clustering

algorithm with a solution of 30 clusters was used to generate the most cohesive clusters with randomly chosen centroids. Generated clusters are shown on line-plot graphs (Figure 26). The background grey lines indicate all 98 significantly regulated miRNAs and the red lines members of specific clusters (Table 2). 5 of these clusters were nearly linearly down-regulated (cluster 4, 5, 10, 13, 19) and 3 up-regulated (cluster 7, 16, 18) across four developmental time points. Some of the clusters' members expression levels (down- or up-regulated) peaked at P5 (cluster 0, 27) or P14 (cluster 9, 27) or the expression was constantly down-regulated across P5, P14 and W6 (cluster 14) and P14, W6 (cluster 2, 29). 56 of these miRNAs were up-regulated at P5 compared to W6 and only 16 miRNAs at W6 vs. P5 (Table 3). Remaining miRNAs were expressed at the similar level at P5 and W6.

Using the Ingenuity Pathway Analysis (IPA) system, we could identified some of the biological functions and diseases in which the miRNAs up-regulated at P5 vs. W6 and W6 vs. P5 are involved (only experimentally validated) (Figure 27). MiRNAs up-regulated at W6 vs. P5 were highly associated with inflammatory response, immunological diseases, respiratory diseases and connective tissue disorders, whereas miRNAs up-regulated at P5 vs.W6 were found to be related with cellular development, tissue development, embryonic development, respiratory system and connective tissue development. Though there was no difference between P5 and W6 when it comes to regulation of bio-functions like cell functionality, basic cell biology and morphology.

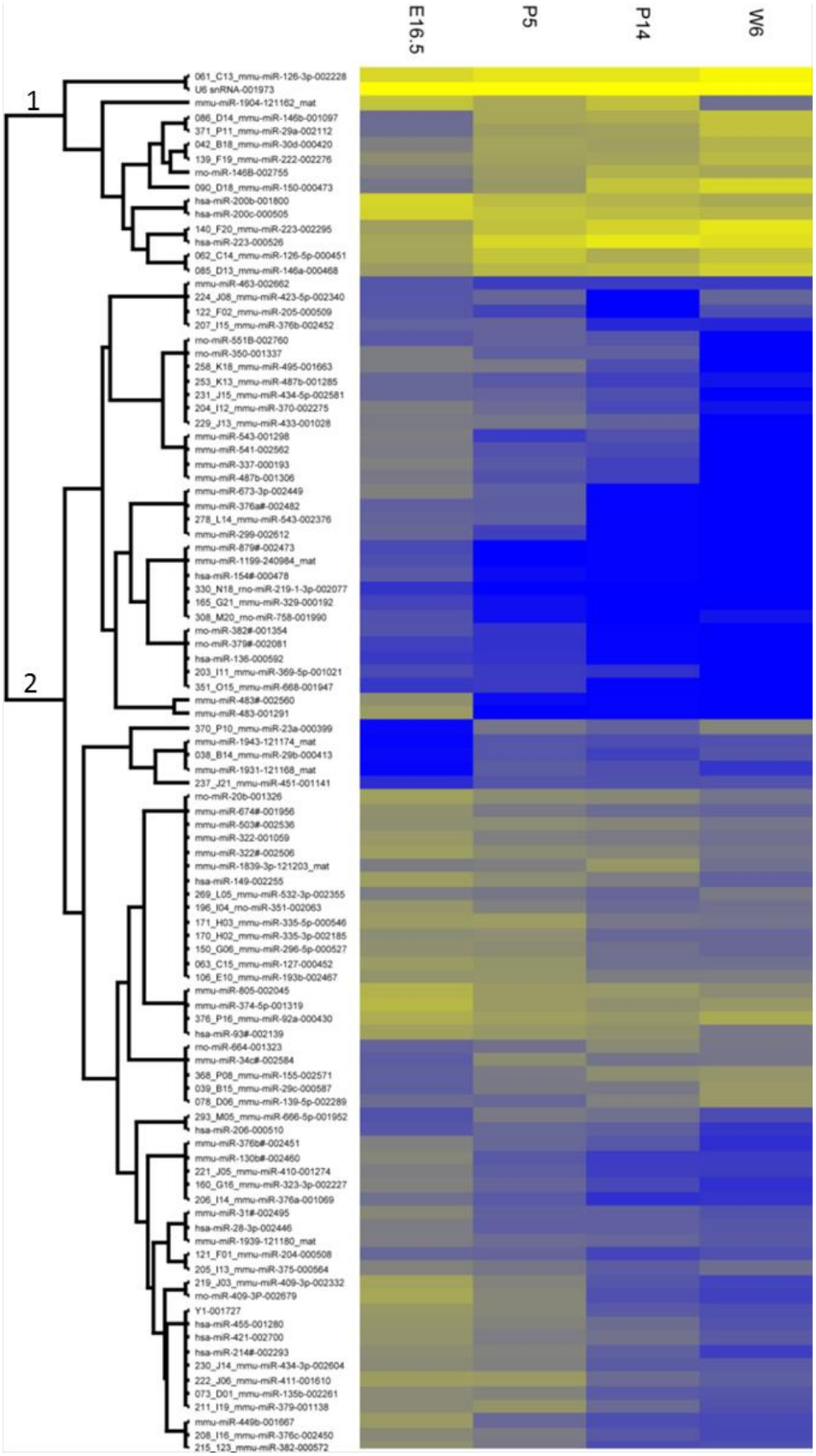

E16.5
P5
P14
W6
1
2
061_C13_mmu-miR-126-3p-002228
U6 snRNA-001973
mmu-miR-1904-121162_mat
086_D14_mmu-miR-146b-001097
371_P11_mmu-miR-29a-002112
042_B18_mmu-miR-30d-000420
139_F19_mmu-miR-222-002276
rno-miR-146B-002755
090_D18_mmu-miR-150-000473
hsa-miR-200b-001800
hsa-miR-200c-000505
140_F20_mmu-miR-223-002295
hsa-miR-223-000526
062_C14_mmu-miR-126-5p-000451
085_D13_mmu-miR-146a-000468
mmu-miR-463-002662
224_J08_mmu-miR-423-5p-002340
122_F02_mmu-miR-205-000509
207_I15_mmu-miR-376b-002452
rno-miR-551B-002760
rno-miR-350-001337
258_K18_mmu-miR-495-001663
253_K13_mmu-miR-487b-001285
231_J15_mmu-miR-434-5p-002581
204_I12_mmu-miR-370-002275
229_J13_mmu-miR-433-001028
mmu-miR-543-001298
mmu-miR-541-002562
mmu-miR-337-000193
mmu-miR-487b-001306
mmu-miR-673-3p-002449
mmu-miR-376a#-002482
278_L14_mmu-miR-543-002376
mmu-miR-299-002612
mmu-miR-879#-002473
mmu-miR-1199-240984_mat
hsa-miR-154#-000478
330_N18_rno-miR-219-1-3p-002077
165_G21_mmu-miR-329-000192
308_M20_rno-miR-758-001990
rno-miR-382#-001354
rno-miR-379#-002081
hsa-miR-136-000592
203_I11_mmu-miR-369-5p-001021
351_O15_mmu-miR-668-001947
mmu-miR-483#-002560
mmu-miR-483-001291
370_P10_mmu-miR-23a-000399
mmu-miR-1943-121174_mat
038_B14_mmu-miR-29b-000413
mmu-miR-1931-121168_mat
237_J21_mmu-miR-451-001141
rno-miR-20b-001326
mmu-miR-674#-001956
mmu-miR-503#-002536
mmu-miR-322-001059
mmu-miR-322#-002506
mmu-miR-1839-3p-121203_mat
hsa-miR-149-002255
269_L05_mmu-miR-532-3p-002355
196_I04_rno-miR-351-002063
171_H03_mmu-miR-335-5p-000546
170_H02_mmu-miR-335-3p-002185
150_G06_mmu-miR-296-5p-000527
063_C15_mmu-miR-127-000452
106_E10_mmu-miR-193b-002467
mmu-miR-805-002045
mmu-miR-374-5p-001319
376_P16_mmu-miR-92a-000430
hsa-miR-93#-002139
rno-miR-664-001323
mmu-miR-34c#-002584
368_P08_mmu-miR-155-002571
039_B15_mmu-miR-29c-000587
078_D06_mmu-miR-139-5p-002289
293_M05_mmu-miR-666-5p-001952
hsa-miR-206-000510
mmu-miR-376b#-002451
mmu-miR-130b#-002460
221_J05_mmu-miR-410-001274
160_G16_mmu-miR-323-3p-002227
206_I14_mmu-miR-376a-001069
mmu-miR-31#-002495
hsa-miR-28-3p-002446
mmu-miR-1939-121180_mat
121_F01_mmu-miR-204-000508
205_I13_mmu-miR-375-000564
219_J03_mmu-miR-409-3p-002332
rno-miR-409-3P-002679
Y1-001727
hsa-miR-455-001280
hsa-miR-421-002700
hsa-miR-214#-002293
230_J14_mmu-miR-434-3p-002604
222_J06_mmu-miR-411-001610
073_D01_mmu-miR-135b-002261
211_I19_mmu-miR-379-001138
mmu-miR-449b-001667
208_I16_mmu-miR-376c-002450
215_123_mmu-miR-382-000572

Figure 25 Expression pattern of miRNAs across different time points of lung development. *Differentially regulated miRNAs across 4 time points are shown on the heat map as median of 3 replicates for each time point; yellow– high expression level, blue– low expression level. Hierarchical clustering indicates similarities of expression level between different miRNAs across the time points. Single miRNAs of selected clusters are listed in Table 2. One-way ANOVA test, $p<0.05$ was applied. Data analysis was performed in collaboration with Dr. Mario Looso from Max Planck Institute in Bad Nauheim. Data was generated and collected at Department of Molecular Pneumology of Philipps University in Marburg (collaborators: Prof. Dr. Bernd Schmeck and Dr. Wilhelm Bertrams). E= embryonic day, P= postnatal day, W= week.*

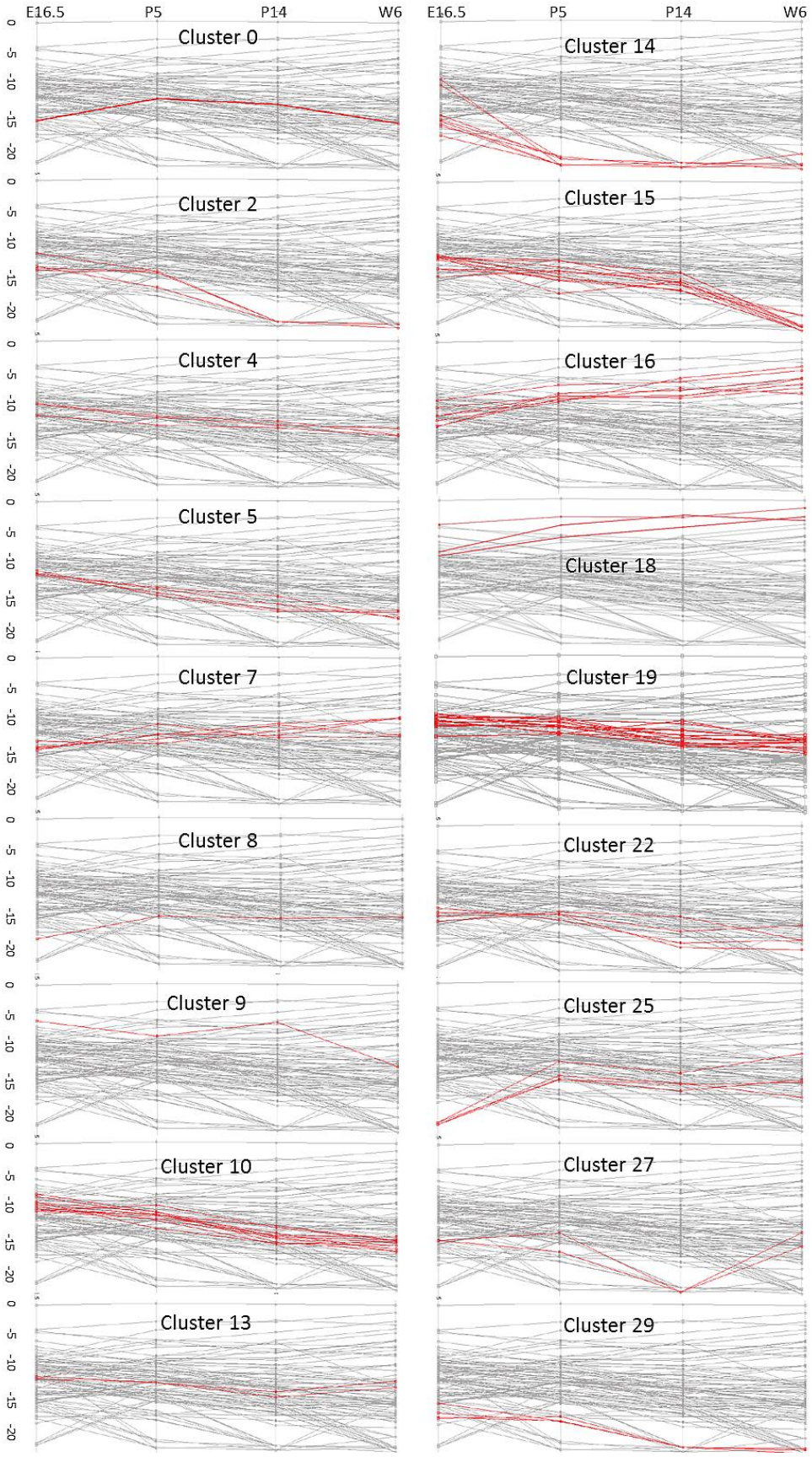

E16.5
P5
P14
W6
Cluster 0
Cluster 2
Cluster 4
Cluster 5
Cluster 7
Cluster 8
Cluster 9
Cluster 10
Cluster 13
Cluster 14
Cluster 15
Cluster 16
Cluster 18
Cluster 19
Cluster 22
Cluster 25
Cluster 27
Cluster 29

Figure 26 Expression pattern of miRNAs within specific clusters. *MiRNAs were grouped according to their expression level similarities across different time points. Grey line plots indicate all 98 significantly regulated miRNAs and red line plots the miRNA members of specific clusters. K-means clustering algorithm was used. Y-axis shows ΔCt values. Data was analyzed in collaboration with Dr. Mario Looso from Max Planck Institute in Bad Nauheim. E= embryonic day, P = postnatal day, W = week after birth.*

Table 2 List of miRNA members within their clusters. P-values are shown. ANOVA test; $p<0.05$.

CLUSTER	miRNAs	p-value	CLUSTER	miRNAs	p-value
Cluster 0	mmu-miR-666-5p	0.0041	**Cluster15**	mmu-miR-370	0.0009
				mmu-miR-433	0.0030
				mmu-miR-434-5p	3.35E-09
				mmu-miR-487b	0.0010
				mmu-miR-495	0.0004
				mmu-miR-337	0.0014
				mmu-miR-487b	3.15E-08
				mmu-miR-541	1.78E-09
				mmu-miR-543	0.0046
				rno-miR-350	0.0021
				rno-miR-551B	0.0094
Cluster2	mmu-miR-543	0.0040	**Cluster16**	mmu-miR-30d	0.0062
	mmu-miR-299	2.67E-07		mmu-miR-146a	1.14E-09
	mmu-miR-376a*	0.0031		mmu-miR-146b	0.0012
	mmu-miR-673-3p	0.0027		mmu-miR-150	3.75E-09
				mmu-miR-222	0.0013
				mmu-miR-29a	0.0001
				rno-miR-146B	0.0004
Cluster 4	hsa-miR-421	5.42E-09	**Cluster 17**	mmu-miR-369-5p	0.0095
	mmu-miR-1939	0.0104			
	mmu-miR-674*	0.0091			
Cluster5	mmu-miR-323-3p	0.0060	**Cluster18**	mmu-miR-126-3p	0.0070
	mmu-miR-410	0.0001		mmu-miR-223	0.0102
	mmu-miR-130b*	0.0012		hsa-miR-200c	0.0104
	mmu-miR-376b*	0.0022		hsa-miR-223	0.0001
Cluster6	mmu-miR-805	0.0135	**Cluster19**	mmu-miR-127	0.0023
				mmu-miR-193b	0.0034
				mmu-miR-296-5p	0.0028
				mmu-miR-335-3p	0.0001
				mmu-miR-335-5p	0.0030
				rno-miR-351	0.0007
				mmu-miR-411	0.0009
				hsa-miR-149	0.0024
				hsa-miR-93*	0.0077
				mmu-miR-1839-3p	0.0002
				mmu-miR-322*	0.0021
				mmu-miR-322	0.0042
				mmu-miR-503*	0.0009
				rno-miR-20b	0.0102
Cluster 7	mmu-miR-29c	0.0009	**Cluster20**	mmu-miR-126-5p	0.0109
	mmu-miR-139-5p	0.0021		mmu-miR-92a	0.0107
	mmu-miR-155	0.0051		hsa-miR-200b	0.0023
	mmu-miR-34c*	0.0028		mmu-miR-374-5p	0.0015

	rno-miR-664	0.0014			
Cluster 8	mmu-miR-451	0.0133	**Cluster 22**	mmu-miR-204 mmu-miR-376a mmu-miR-376b hsa-miR-206	0.0070 0.0116 0.0050 0.0019
Cluster 9	mmu-miR-1904	0.0003	**Cluster24**	mmu-miR-463	0.0060
Cluster 10	mmu-miR-135b mmu-miR-376c mmu-miR-379 mmu-miR-382 mmu-miR-409-3p mmu-miR-434-3p hsa-miR-214* hsa-miR-455 mmu-miR-449b rno-miR-409-3P Y1	0.0019 0.0020 0.0116 0.0014 0.0006 0.0014 0.0007 0.0047 0.0017 6.50E-06 0.0072	**Cluster25**	mmu-miR-29b mmu-miR-23a mmu-miR-1931 mmu-miR-1943	0.0068 0.0099 0.0007 0.0002
Cluster13	mmu-miR-375 mmu-miR-532-3p	0.0032 0.0099	**Cluster27**	mmu-miR-205 mmu-miR-423-5p	0.0082 0.0003
Cluster14	mmu-miR-329 rno-miR-758 rno-miR-219-1-3p hsa-miR-154* mmu-miR-1199 mmu-miR-483* mmu-miR-483 mmu-miR-879*	0.0063 0.0067 0.0006 0.0009 8.79E-06 0.0015 7.62E-07 6.24E-05	**Cluster 28**	hsa-miR-28-3p mmu-miR-31*	0.0108 0.0049
			Cluster 29	mmu-miR-668 hsa-miR-136 rno-miR-379* rno-miR-382*	0.0072 0.0133 0.0079 0.0058

Table 3 MiRNAs up-regulated at P5 vs. W6 and W6 vs. P5.

Up-regulated at P5 vs. W6			**Up-regulated at W6 vs. P5**
rno-miR-350	mmu-miR-376a	rno-miR-351	mmu-miR-30d
rno-miR-551B	hsa-miR-206	mmu-miR-411	mmu-miR-146a
hsa-miR-421	mmu-miR-1904	hsa-miR-149	mmu-miR-146b
mmu-miR-1939	mmu-miR-135b	hsa-miR-93*	mmu-miR-150
mmu-miR-674*	mmu-miR-376c	mmu-miR-1839-3p	mmu-miR-222
mmu-miR-323-3p	mmu-miR-379	mmu-miR-322*	mmu-miR-29a
mmu-miR-410	mmu-miR-382	mmu-miR-322	rno-miR-146B
mmu-miR-130b*	mmu-miR-409-3p	mmu-miR-503*	mmu-miR-126-3p
mmu-miR-376b	mmu-miR-434-3p	rno-miR-20b	mmu-miR-223
mmu-miR-668	hsa-miR-214*	mmu-miR-433	hsa-miR-200c
hsa-miR-136	hsa-miR-455	mmu-miR-434-5p	hsa-miR-223
rno-miR-379*	mmu-miR-449b	mmu-miR-487b	mmu-miR-29c
rno-miR-382*	rno-miR-409-3P	mmu-miR-495	mmu-miR-139-5p
mmu-miR-543	mmu-miR-127	mmu-miR-337	mmu-miR-155
mmu-miR-299	mmu-miR-193b	mmu-miR-487b	mmu-miR-34c*
mmu-miR-376a*	mmu-miR-296-5p	mmu-miR-541	rno-miR-664
mmu-miR-673-3p	mmu-miR-335-3p	mmu-miR-543	
mmu-miR-370	mmu-miR-335-5p	mmu-miR-204	
		mmu-miR-369-5p	

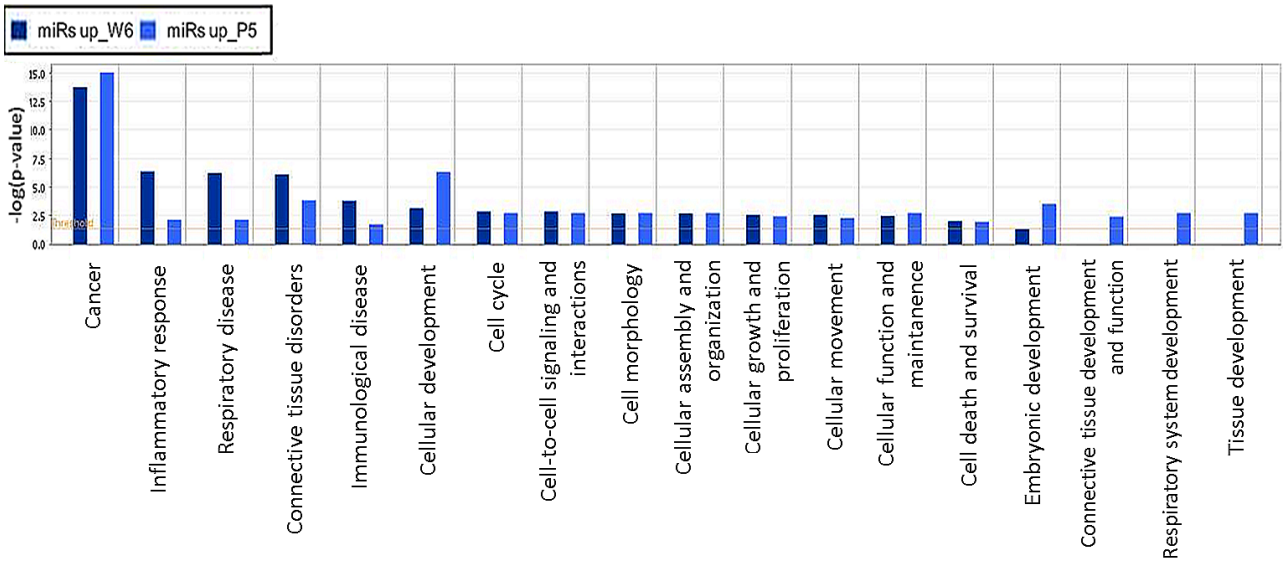


Figure 27 Bio-functions of miRNAs up-regulated at P5 vs. W6 and W6 vs. P5 in pulmonary PDGFRα-positive cells. *Fisher's exact test was used, threshold value was set as $p<0.05$. Data were analyzed through the use of IPA (Ingenuity® Systems, www.ingenuity.com).*

5.7.1. *In silico* analysis of miRNAs putatively targeting *Pdgfrα* 3'-UTR

Using TargetScan algorithm, we identified only 4 from the 98 significantly regulated miRNAs that were predicted to target the *Pdgfrα* gene along the highly conserved binding sites for vertebrates, whereas 20 miRNAs were associated with poorly conserved binding sites among vertebrates (Table 4). These 4 identified miRNAs belong to clusters: 4 (hsa-miR-421), 10 (mmu-miR-449b) and 19 (hsa-miR-149, rno-miR-20b); all of them were linearly down-regulated across the four time points (E16.5=>W6). Using miRTarBase, we identified no miRNAs (within 98 significantly regulated) which were experimentally validated to target *Pdgfrα*. MiR-20b-5p, the member of miR-17-5p family appeared to have the highest probability of preferential site conservation within *Pdgfrα* gene, moreover it was found to repress *Pparγ* and *Tgfbr 2* (validated targets).

Table 4 MiRNAs predicted to bind to highly or poorly conserved sites within the *Pdgfrα* 3'-UTR gene fragment. MiRNAs were identified by the use of TargetScan Release 5.2. Affiliation of miRNAs to clusters according to K-means clustering is shown. C = Cluster.

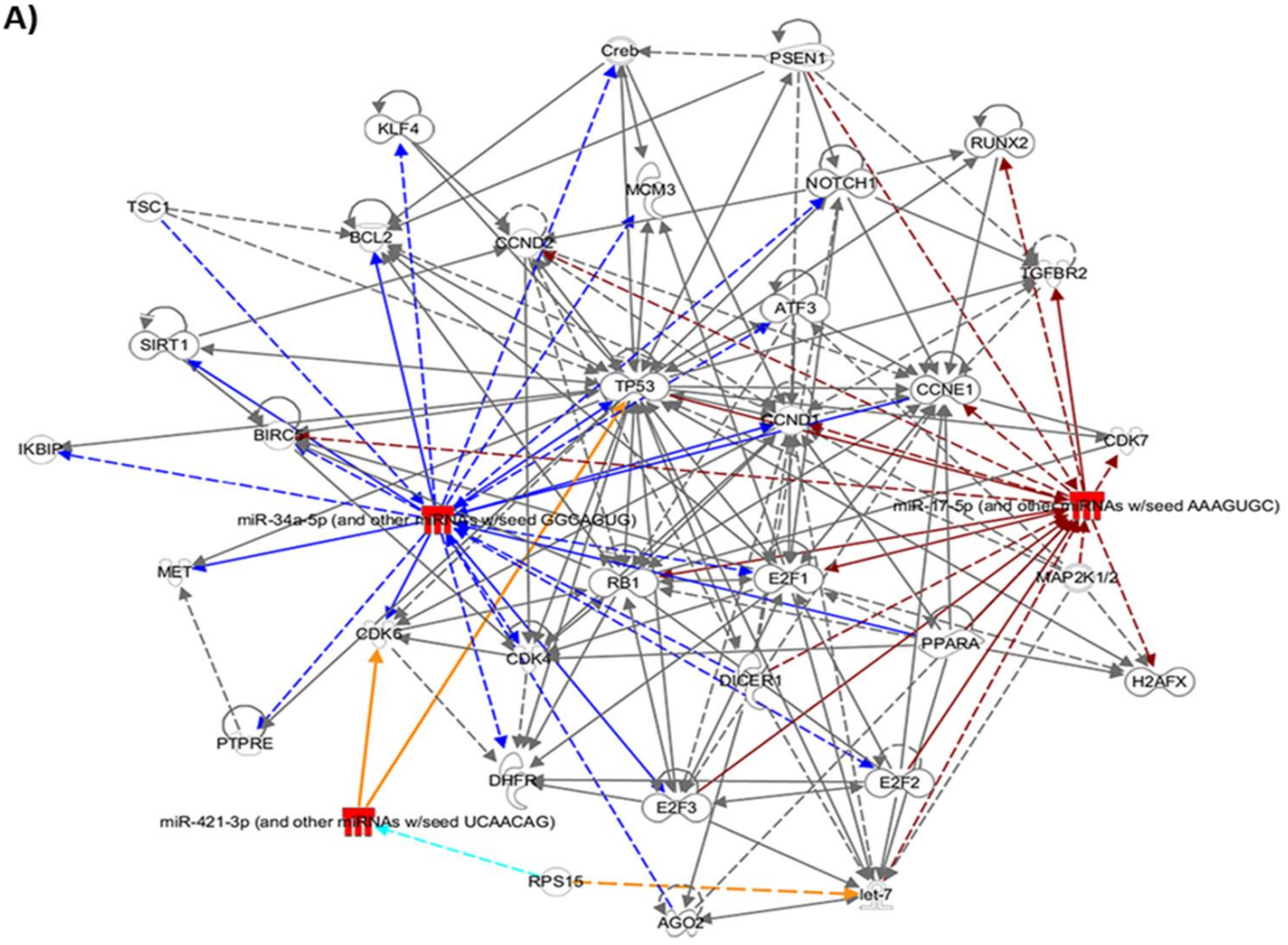
miRNAs binding to highly conserved sites within the *Pdgfrα* 3'-UTR gene fragment	**miRNAs binding to poorly conserved sites within the *Pdgfrα* 3'-UTR gene fragment**
rno-miR-20b-5p (C19; miR-17-5p family) (binding to sites with higher probability of preferential conservation) hsa-miR-149-5p (C19) (conserved only among mammals) hsa-miR-421 (C4) (conserved only among mammals) mmu-miR-449b (C10; mir-34a-5p family)	mmu-miR-146b (C16), mmu-miR-29a (C16), mmu-miR-30d (C16), mmu-miR-150 (C16), mmu-miR-223 (C18), mmu-miR-146a (C16), hsa-miR-223 (C18), mmu-miR-205 (C27), mmu-miR-495 (C15), mmu-miR-370 (C15), mmu-miR-433 (C15), rno-miR-758 (C14) hsa-miR-136 (C29), mmu-miR-29b (C25), mmu-miR-335-5p (C19), mmu-miR-374-5p (C20), mmu-miR-29c (C7), mmu-miR-410 (C5), hsa-miR-421 (C4), mmu-miR-376c (C10)

MiRNAs binding to the highly or poorly conserved sites on the *Pdgfrα* 3'-UTR gene fragment were further analyzed using Ingenuity Pathway Analysis (IPA). Experimentally known or proposed interactions between miRNAs and other molecules were generated (Figure 28 A, B). Each miRNA was assigned into and represented as miRNA family according to their "seed region" sequence similarity (Appendix, Table A1). The network created by miRNAs binding to highly conserved sites within the *Pdgfrα* 3'-UTR gene fragment consisted of 2 sub-networks (connected by orange lines on Figure 28). They included: 12 transcription regulators (Klf4, Notch1, Runx2, Sirt1, Atf3, Tp53, Ccne1, Rb1, E2f1, H2afx, E2f2, E2f3), 1 peptidase (Psen1), 3 enzymes (Mcm3, Dicer1, Dhfr), 1 transporter (Bcl2), 5 growth factors (Tgfbr2, Cdk7, Met, Cdk5, Cdk4), 1 translation regulator (Ago2), 1 phosphatase (Ptpre), 1 pre-mir (let-7), 1 ligand-dependent nuclear receptor (Ppara), and 8 other molecules (Creb, Tsc1, Ccnd2, Ccnd1, Ikbip, Birc5, Map2k1/2, Rps15). The network formed by miRNAs binding to the poorly conserved sites within the *Pdgfrα* 3'-UTR gene fragment included 3 sub-networks (merged by orange lines on Figure 28). They consisted of: 2 transcription regulators (Smad7, Tp53), 1 phosphatase (Inppl1), 2 kinases (Mapk14, Chuk), 3 growth factors/ cytokines (Il-8, Ctgf, Bdnf), 1 enzyme (Dcp2), 1 ligand-dependent nuclear factor (Nr0b2), 1 translation regulator (Ago2), 6 other molecules (Cofilin, F-actin, Cg, Fsh, Rps15, Mapre2), 7 pre-miRNAs (let-7, pre-mir-146, -335, -23, -17, -30, -10), and other mature miRs families (miR-20a-3p, -23a-3p,

let-7a-5p). The full nomenclatures of these molecules are shown in the Appendix, Table A2.

The bio-functions of these miRNAs were also analyzed. MiRNAs binding to the poorly conserved sites of the *Pdgfrα* 3'-UTR gene fragment were associated with cancer, inflammatory/infectious diseases, inflammatory response, connective tissue disorder and respiratory diseases. MiRNAs binding to the highly conserved sites within the *Pdgfrα* 3'-UTR gene fragment were favored to regulate cellular function, cell-to-cell interaction and signaling, cell cycle and morphology, cellular growth, proliferation and movement, cell death and survival, connective tissue and respiratory system development, embryonic development, and tissue development and morphology (Figure 29).

A)



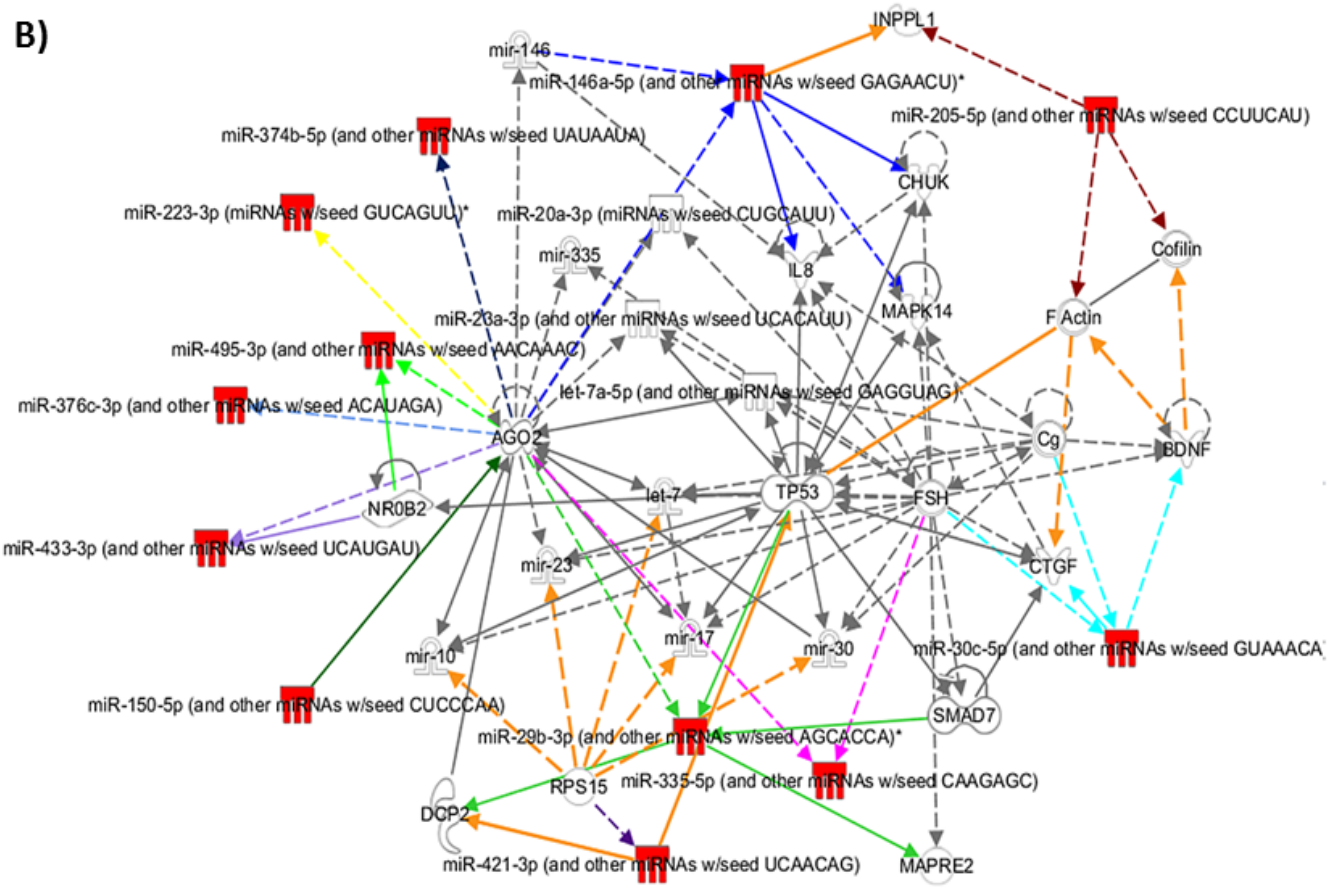


Figure 28 Networks of miRNA/molecule interactions associated with A) miRNA binding to the highly or B) poorly conserved sites within Pdgfrα 3'-UTR fragment. *Interactions between specific miRNAs and other molecules are shown as differentially colored lines. Known interactions are indicated by solid lines, proposed interactions by dashed lines. MiRNAs discovered in the present study to be regulated during lung development in the PDGFRα-positive cells are presented as red. Orange lines show the connection between sub-networks. Each MiRNA is labeled as the representative of its family with the same "seed region" (see Table 4 and Appendix, Table A1). Hsa-miR-149-5p was not experimentally validated to interact with any molecules (A). Data were analyzed through the use of IPA (Ingenuity® Systems, www.ingenuity.com).*

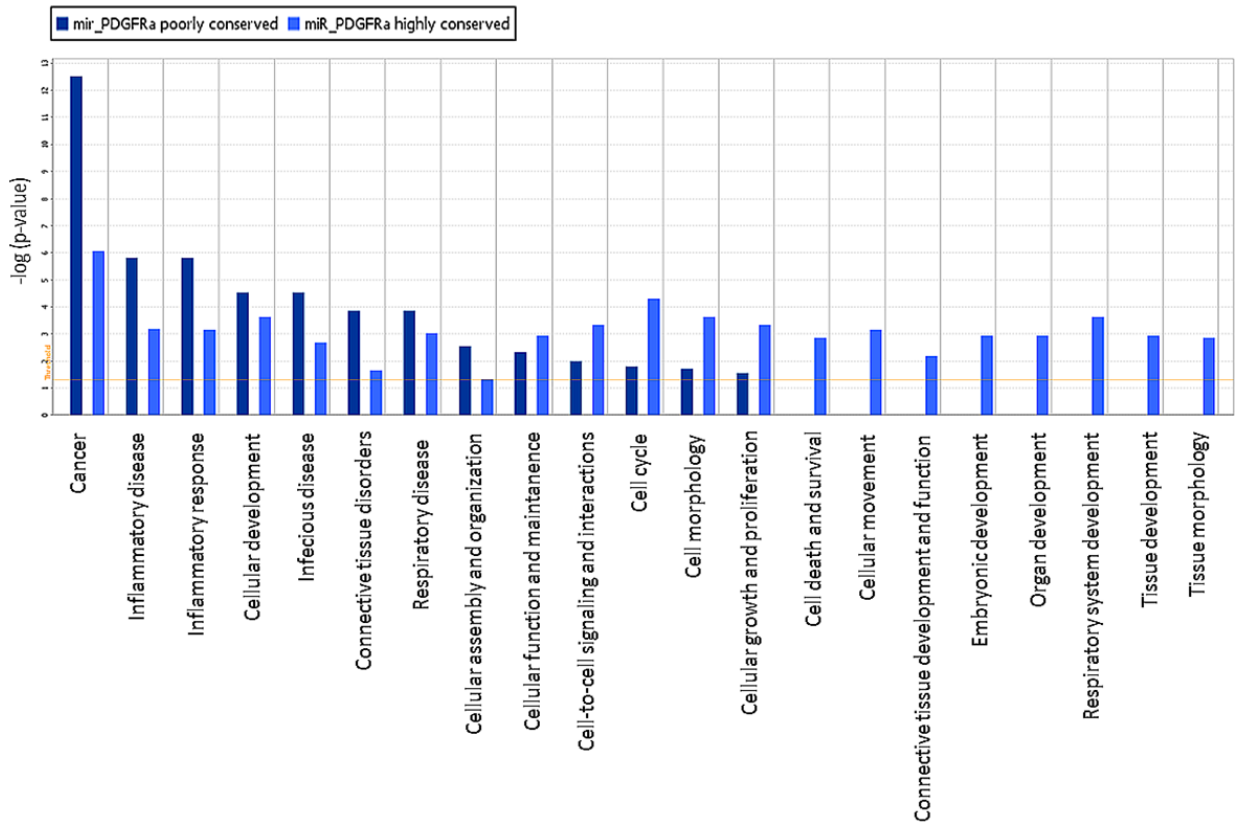


Figure 29 Bio-functions of miRNAs putatively targeting Pdgfrα. *MiRNAs were grouped according to highly or poorly conserved sites putatively recognized on Pdgfrα 3'-UTR target. Fisher's exact test was used. Threshold value was set as $p<0.05$. Data were analyzed through the use of IPA (Ingenuity® Systems, www.ingenuity.com).*

5.8. Next-generation sequencing of PDGFRα-positive cells transcriptome

All next-generation sequencing procedures were performed by Dr. Stefan Günther from Max Planck Institute in Bad Nauheim. Analysis of the data was done in collaboration with biostatisticians from Max Planck Institute in Bad Nauheim.

5.8.1. Quality of generated cDNA libraries

cDNA libraries were generated from RNA isolated from postnatal day 5 (P5), week 6 (W6), and negative control (NC; 1:1 cell from P5 and W6 lung homogenate minus CD45- and PDGFRα- positive cells). Total RNA integrity number (RIN) ranged approximately from 7-8.5 (Figure 30B). rRNA was depleted and transcriptome RNA concentrated (Figure 30A). Each library was prepared in duplicate (see section 4.12, Figure 12). No size selection of cDNA libraries was performed due to a homogeneous size distribution between samples (Figure 31). Samples were then subjected to next-generation sequencing.

5.8.2. Sequencing data quality and depth of coverage

Generated sequencing data were analyzed for the quality and depth of coverage of the raw data before trimming. Number of sample bases, Q20 bases, number of reads, mean read length was included (Figure 32). The raw data were trimmed to get rid of poor quality reads and increase the quality of results. The acceptable threshold was determined for all samples during the validation process: min sequence length - 18bp, max sequence length - 150bp (Table 5). Distribution of sequence length over all sequences before and after trimming is shown (Figure 33). Sequence reads were mapped basing on a reference genome (*mus musculus*) by use of TopHat/Bowtie2; percentage of mapped reads is shown (Table 6). *De novo* and bad status transcripts were excluded. Replicates (s1/ s2, s3/ s4, s5/ s6) concordance measured by Pearson correlation was approximately 93-95% showing good reproducibility (Figure 34). Data were further analyzed for differentially regulated genes.

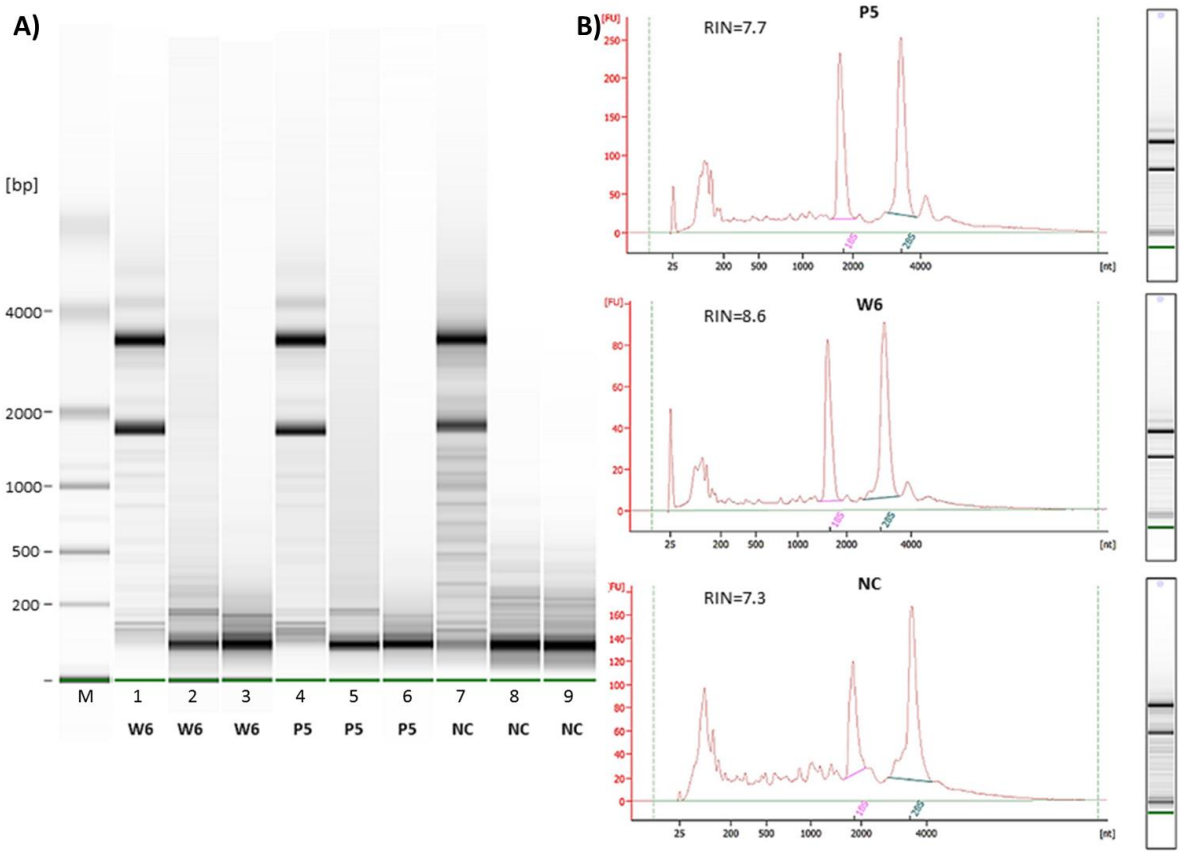


Figure 30 Total RNA integrity and rRNA-depleted transcriptome RNA. *Eukaryote total RNA Pico Kit was used. W6 = week 6, P5 = postnatal day 5, NC = negative control.* ***A)*** *Gel image of total RNA trace (line 1, 4, 7) and RNA after rRNA depletion (line 2, 3, 5, 6, 8, 9) are shown. M = marker of RNA size.* ***B)*** *Histograms show total RNA trace with the 28S and 18S rRNA peaks. RNA integrity number (RIN) is shown. Y-axis shows fluorescence units (FU), x-axis shows number of nucleotides [nt].*

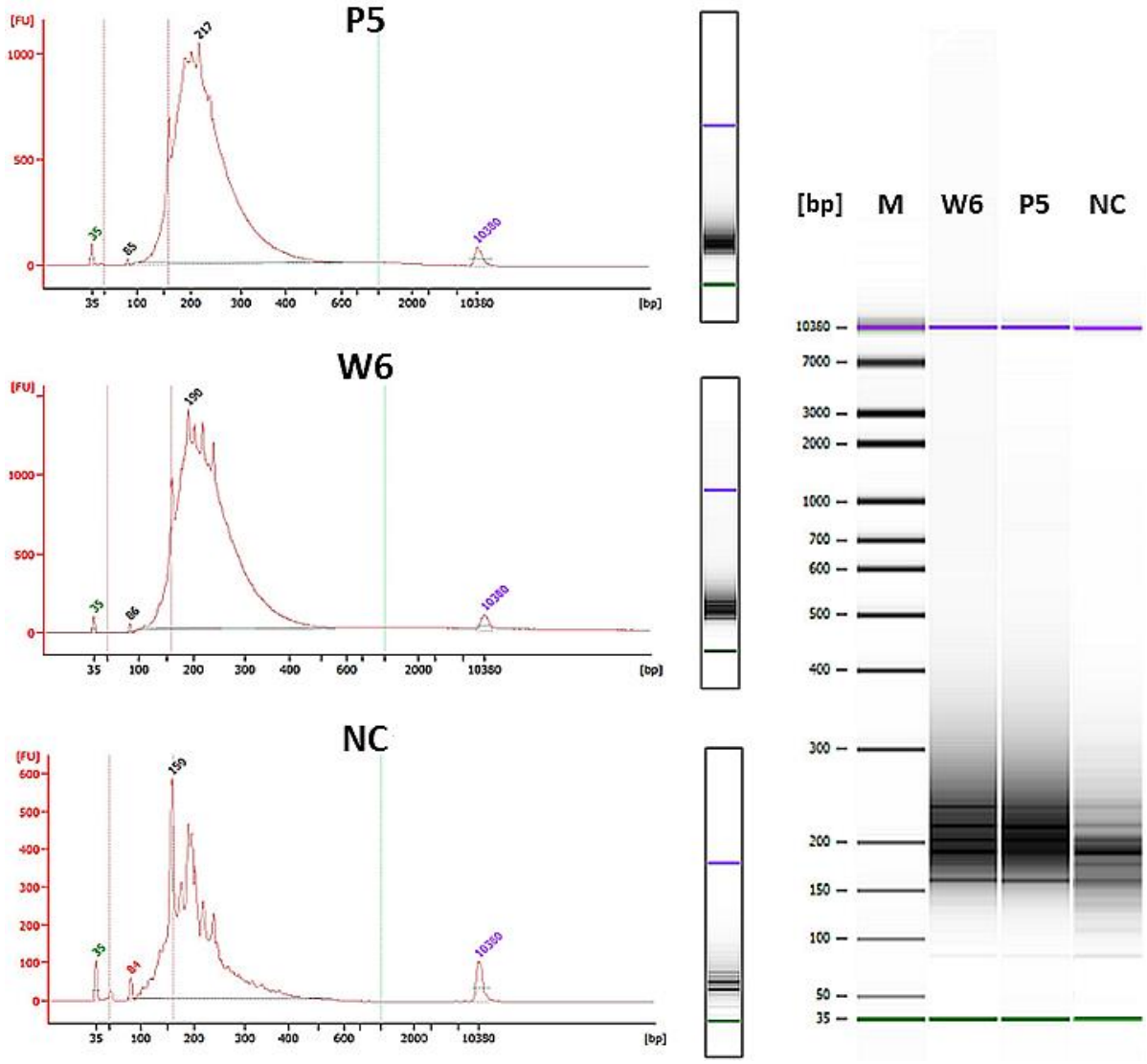


Figure 31 cDNA libraries generated from RNA isolated from lung PDGFRα-positive cells and negative control. *Histograms show size distribution of cDNA libraries; x-axis shows length of DNA fragments in bp; y-axis shows fluorescence units (FU). Gel image for each cDNA library is shown. M = marker, W6 = week 6, P5 = postnatal day 5, NC = negative control. Green or purple bands indicate max or min size marker.*

	Sample Bases	Q20Bases (top quality reads)	Number of Reads	Mean read length	
Lib I	4,249,833,186	3,544,369,793	41,304,620	102 bp	
Lib V	2,298,256,128	1,945,657,420	25,261,439	90 bp	
Lib III	3,523,818,209	2,877,228,540	34,389,544	102 bp	
Lib VI	3,385,007,164	2,758,780,055	36,855,392	91 bp	
Lib II	3,275,891,273	2,668,947,247	37,109,597	88 bp	
Lib IV	2,678,220,123	2,145,294,900	30,258,826	88 bp	

Figure 32 Sequencing data quality and depth of coverage (raw data before trimming). *Number of sample bases, Q20 bases, number of reads and mean read length are shown. Histograms show distribution of reads for each library [bp]. LibI/ II= W6, LibIII/ IV= P5, LibV/ VI= NC. Lib = library, NC= negative control, W6 = week 6, P5 = postnatal day 5.*

Table 5 Trimmed sequencing data: min length - 18bp, max length - 150bp. Number of reads (count) and sample bases (sum) are shown. Lib = library, NC = negative control.

	LibI_W6	**LibII_W6**	**LibIII_P5**	**LibIV_P5**	**LibV_NC**	**LibVI _NC**
Count	34,136,032	27,728,587	27,521,434	21,511,785	19,967,485	27,833,076
Sum	3,610,088,275	2,599,105,385	2,851,259,223	2,072,509,720	1,958,138,812	2,702,654,368
Mean Length (bp)	106	94	104	96	98	97

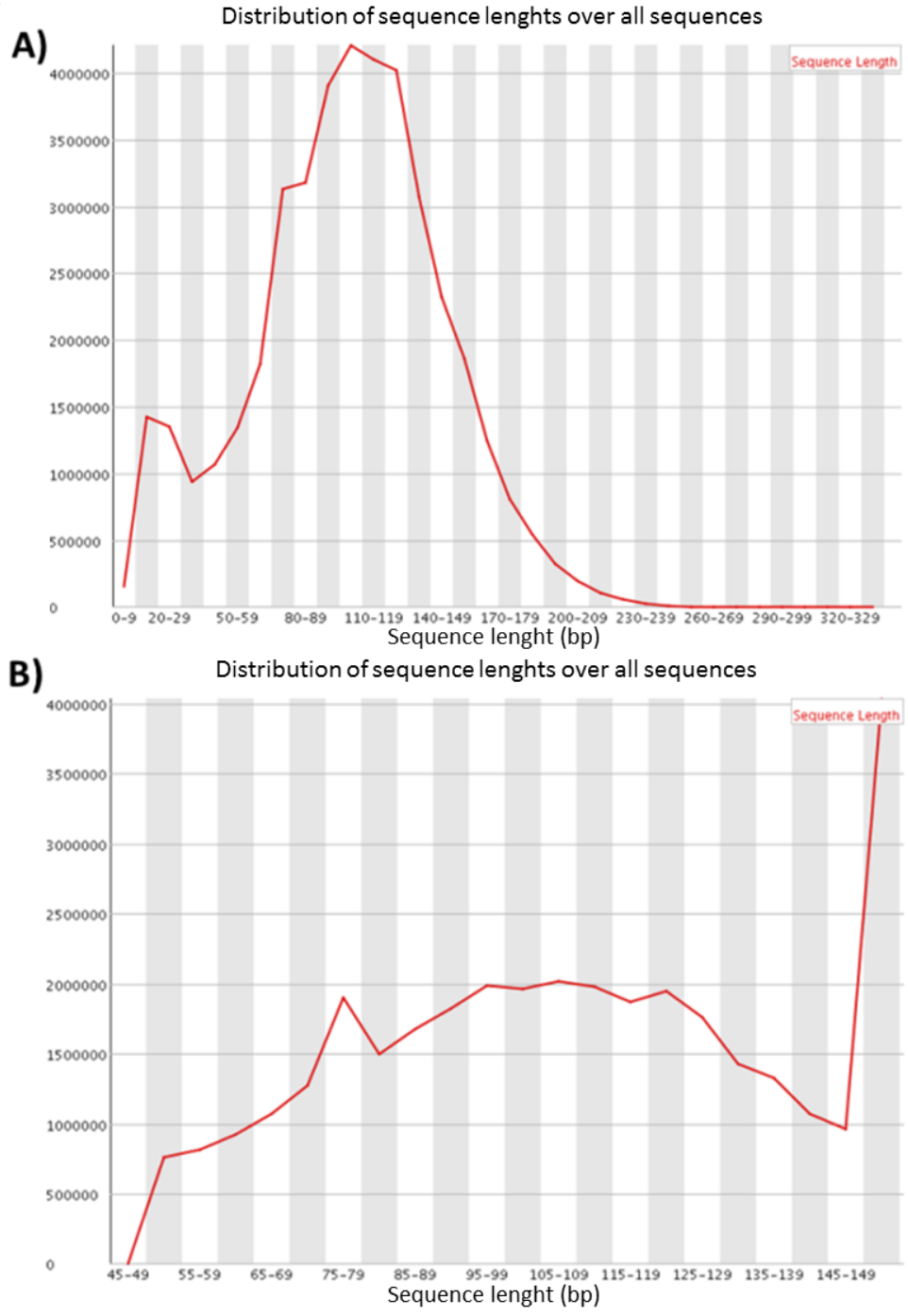


Figure 33 Distribution of sequence length over all sequences. *Y-axis shows number of reads.* ***A)*** *Raw data,* ***B)*** *trimmed data.*

Table 6 Mapped reads referred to the *mus musculus* genome. TopHat/Bowtie2 after data trimming was used; *de novo* transcripts and bad status were excluded. Percentage and number of mapped reads are shown. Lib = library, NC = negative control.

	Lib I_w6	Lib II_w6	Lib III_p5	Lib IV_p5	Lib V_ NC	Lib VI_ NC
Mapped	20240239 (59.3%)	14205194 (51.2%)	13835536 (50.3%)	9633423 (44.8%)	13562105 (67.9%)	18236745 (65.5%)

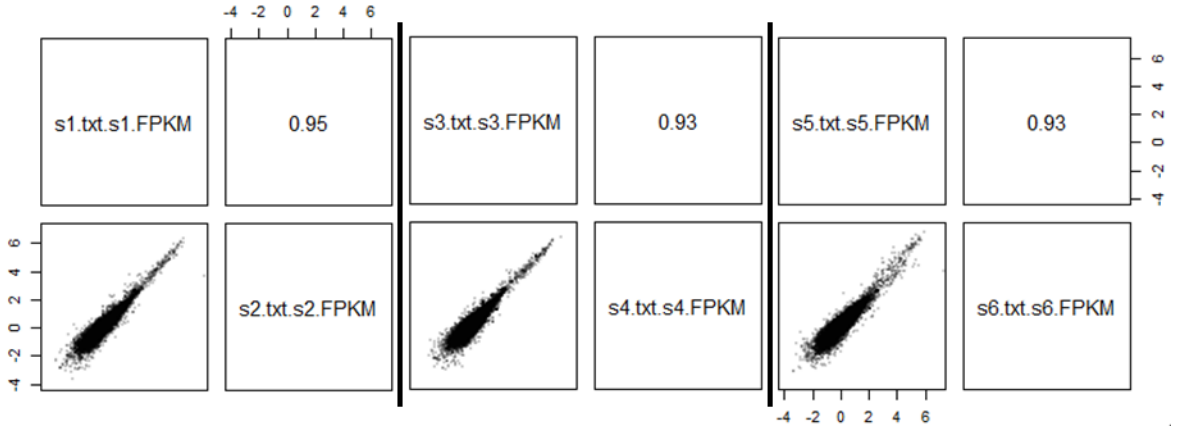


Figure 34 Pairwise comparison of global transcripts expression level in sample replicates. *The x- and y-axes are plotted on a log10-scaled FPKM and show transcripts expression of P5, W6 and NC replicates. Pearson correlation is shown. Replicates: s1 W6/ s2 W6; s3 P5/ s4 P5; s5 NC/ s6 NC. FPKM = Fragments Per Kilobase of transcript per Million mapped reads, P5 = postnatal day 5, W6 = week 6, NC = negative control.*

5.8.3. Differentially regulated genes in PDGFRα-positive cells and control lung cells at P5 and W6

Sequencing data from PDGFRα-positive and negative control lung cells were analyzed according to differentially expressed genes at postnatal day 5 (P5) and week 6 (W6). Changes of gene expression levels (FPKM values) in different samples were presented as log2 (fold change); fold change>2, q<0.05. Number of differentially regulated genes was compared between W6 vs. P5, W6 vs. NC and P5 vs. NC (Figure 35A). Venn diagrams show 137 overlapping genes common for all three relationships, 909+137 common for W6 vs. P5 and W6 vs. NC, 1170+137 common for W6 vs. P5 and P5 vs. NC, and 740+137 common for W6 vs. NC and P5 vs. NC (Figure 35A and B). Number of up- and down-regulated genes among each relationship is also shown (Figure 35C). 394 genes were up-regulated at both P5 and W6 compering to NC from which 75 were differentially regulated between P5 and W6 (Figure 35D). Only 6 of these 75 genes were up-regulated at P5 vs. W6 and remaining 69 at W6 vs. P5 (Appendix, Table A4). Based on criteria of

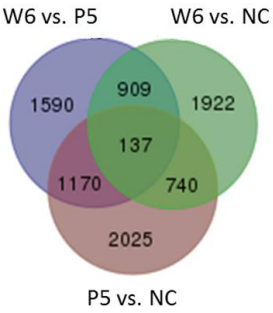
fibroblast structure, regulation of cell differentiation or septa formation, differentially expressed genes were selected and compared between W6 vs. P5, W6 vs. NC and P5 vs. NC (Figure 36); log2 (FC) values of these genes are shown in Appendix, Table A3. Genes up-regulated at P5 vs. W6 (Figure 36A) were as follows: alpha smooth muscle actin (*Acta2*), elastin (*Eln*), myosin regulatory light polypeptide 7 and 9 (*Myl7, 9*), myosin heavy chain 11 (*Myh11*), smoothened (*Smo*), solute carrier family 27A6 (*Slc27a6*), sprouty 1 and 4 (*Spry1,4*), types of collagen (*Col*), connective tissue growth factor (*Ctgf*), transforming growth factor beta 3 and 2 (*Tgfb3, 2*), transforming growth factor beta-induced (*Tgfbi*), transforming growth factor beta receptor 3 (*Tgfbr3*), and latent transforming growth factor beta binding protein 2, 3, 4 (*Ltbp2, 3, 4*). Genes down-regulated at P5 compared to W6 were: coactosin-like protein (*Cotl1*), interleukin 17 receptor a (*Il17ra*), retinoid X receptor alpha (*Rxra*), peroxisome proliferator-activated receptor gamma (*Pparγ*), thymocyte differentiation antigen 1 (*Thy1*; CD90), lipoprotein lipase (*Lpl*), transforming growth factor beta receptor 1 and 2 (*Tgfbr1, 2*) and *mir-29a*. Platelet-derived growth factor receptor alpha (*Pdgfrα*) and platelet-derived growth factor receptor-like (*Pdgfrl*) were up-regulated at P5 vs. W6 [log2 (2.72) and log2 (3.85), respectively]. Additionally, at W6 vs. NC most of the genes were down-regulated at W6 with the exception of: *Cotl1*, *Il17ra*, *Pparγ*, *Lpl*, *Thy1*, *Tgfbi*, *Tgfbr1*, *mir-29a* and *mir-24-2* (Figure 36B). *Pdgfrα* was down-regulated at W6 vs. NC, whereas the expression at P5 was not significantly different compared to NC. Genes like *Acta2*, *Myh11*, *Myl7*, *Cotl1*, *Slc27a6*, *Il17ra*, *Pdgfrl*, *Tgfbi*, *Ltbp2*, *mir-24-2* and another isoform of *Col1a2* (this gene was not expressed at W6 and in NC) were up-regulated at P5 vs. NC (Figure 36C). Two miRNA genes were specifically expressed at P5 (significantly lower at W6 and not expressed in NC cells), i.e.: *mir-382* and *mir-487b* (Figure 36C).

5.8.4. Gene set enrichment analysis of RNA-Seq dataset from PDGFRα-positive cells

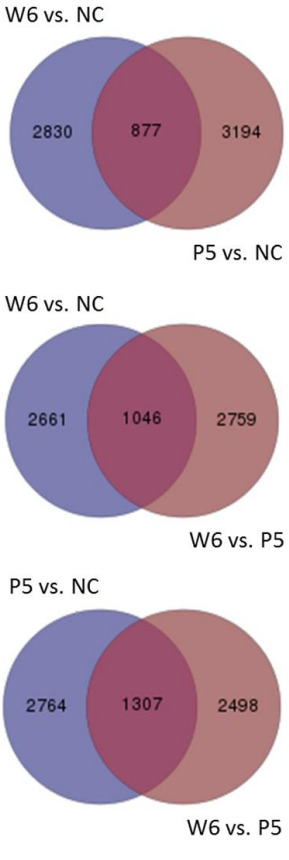
Gene set enrichment analysis (GSEA) was used to analyze differentially expressed gene sets in PDGFRα-positive cells at the time points P5 and W6. Two categories of gene sets were analyzed: gene sets involved in canonical pathways and gene sets sharing 3'-UTR microRNA binding motifs. Gene sets were considered as significantly different with a

FWER (Family-wise error rate) <0.05. The global overview on tested gene sets illustrates the total number of tested gene sets and their enrichment scores as well as significance and normalized enrichment scores (Figure 37A). For miRNA binding motifs we did not find a significantly different gene set for FWER<0.05 and FDR<0.25 at both P5 and W6 time points (Figure 37A, lower graph). Only 1 gene set (genes targeted by miR-155) was significantly enriched for p-value<0.05 at W6, while 9 gene sets (miR-137, -218, -522, -196A, B, -422A, B, -489, -10A, B, -374, -384) were significantly enriched at P5. When it comes to gene sets related with canonical pathways, we found 13 gene sets to be enriched (FWER<0.05) at P5 (Figure 37B) and 70 at W6 (Figure 37C shows top 13 gene sets; Appendix, Table A5, remaining 57). Gene sets involved in ECM organization, collagen formation, muscle contraction, NCAM interactions, integrin1 pathway, cardiomyopathy, WNT and Hedgehog pathways signaling were significant for P5. Gene sets related with immunological system and immune system response regulation, infectious diseases, regulation of translation, and mRNA activation and metabolism were significant for W6.

A)



B)



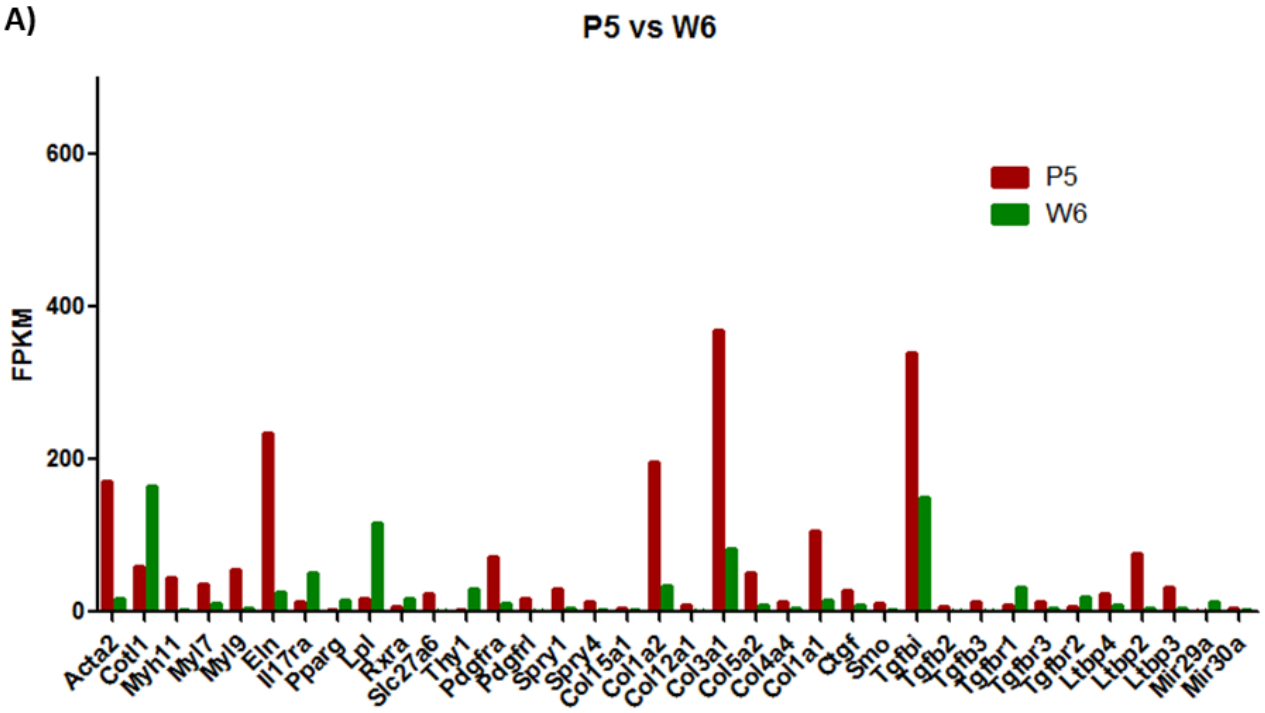
C)

w6 vs. p5	
w6 > p5	1562
w6 < p5	2243
w6 vs. control	
w6 > control	2492
w6 < control	1215
p5 vs. control	
p5 > control	3255
p5 < control	816

D)

w6/p5 commonly up vs. control	**394**
of those: regulated in w6 vs. p5	**75**

Figure 35 Number of differentially regulated genes of PDGFRα-positive lung cells between W6 vs. P5, W6 vs. NC and P5 vs. NC. *Fold change (FC) cut-offs of >2 and significance for q-values of <0.05 were applied.* ***A)*** *Venn diagram of commonly differentially expressed genes among W6 vs. P5, W6 vs. NC and P5 vs. NC.* ***B)*** *Venn diagram of commonly differentially expressed genes between separated two groups, W6 vs. NC and P5 vs. NC, W6 vs. NC and W6 vs. P5, P5 vs. NC and W6 vs. P5.* ***C)*** *Number of up- and down-regulated genes within each group.* ***D)*** *Genes up-regulated at both W6 and P5 compering to NC, and those of them differentially regulated between W6 and P5. NC = negative control, W6 = week 6 after birth, P5 = postnatal day 5.*



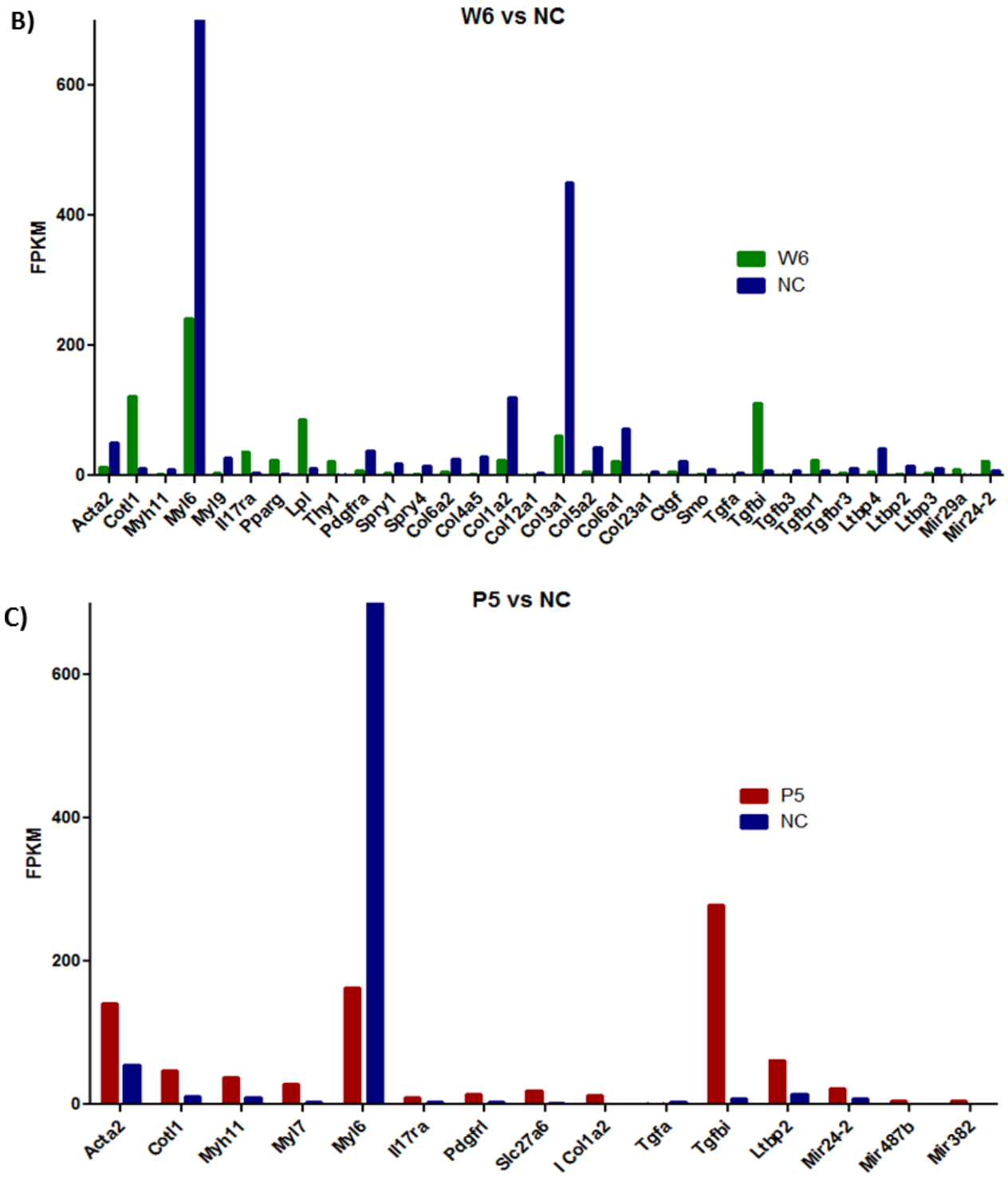
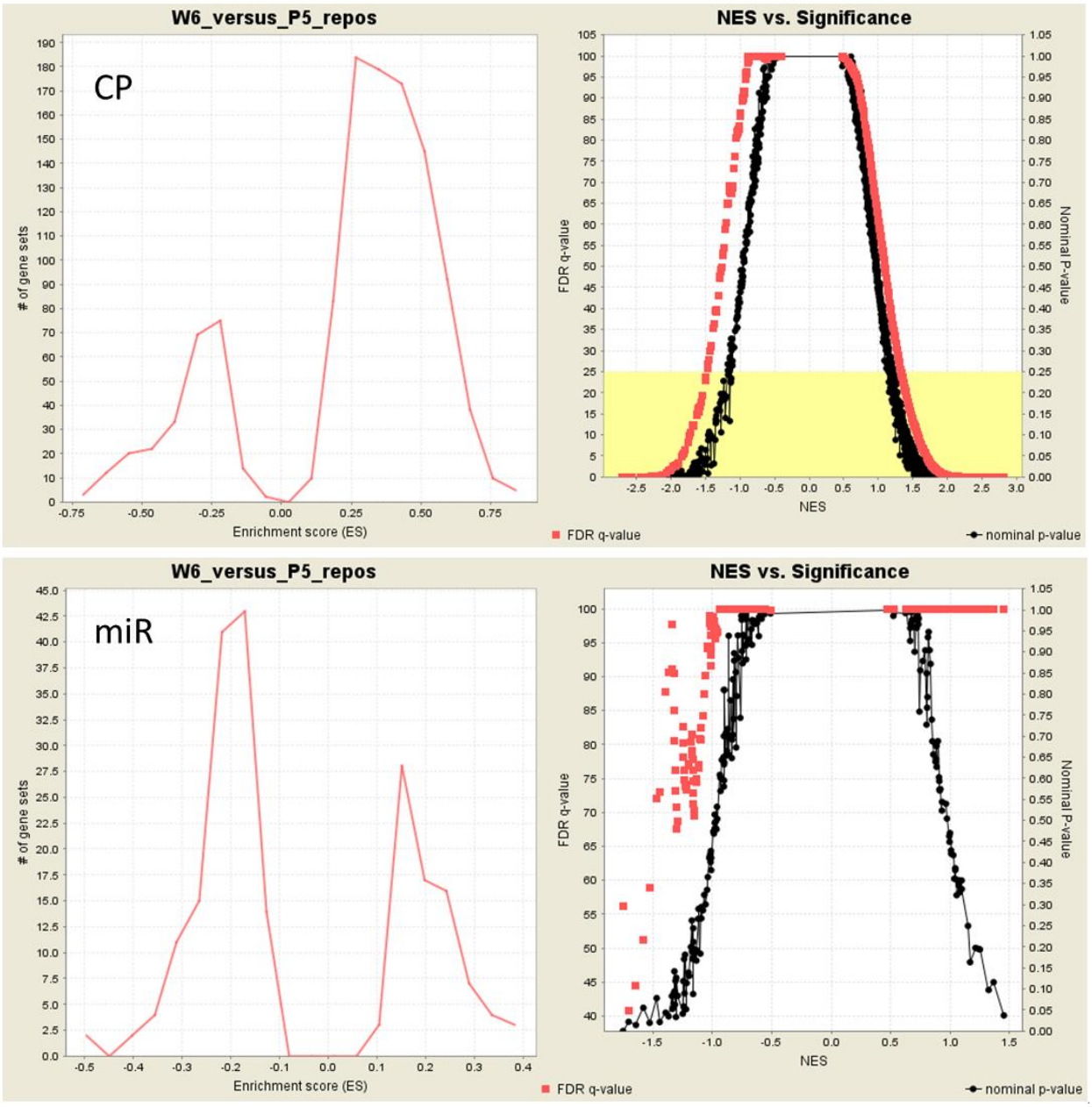


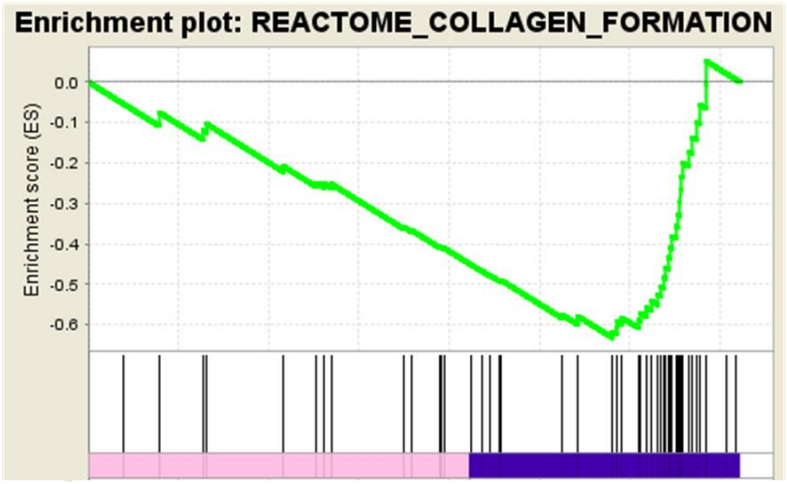
Figure 36 Selected differentially regulated genes of PDGFRα-positive lung cells at P5, W6 and in NC. *Genes were selected based on specific functions during lung development (see text above). Y-axes show FPKM values as the transcripts abundance of specific genes differentially expressed between* ***A)*** *P5 vs. W6,* ***B)*** *W6 vs. NC and* ***C)*** *P5 vs. NC. q-value (FDR)<0.05, FC>2 cut-offs were*

applied. FC = fold change, NC = negative control, W6 = week 6 after birth, P5 = postnatal day 5, FPKM = Fragments Per Kilobase of transcript per Million.

A)

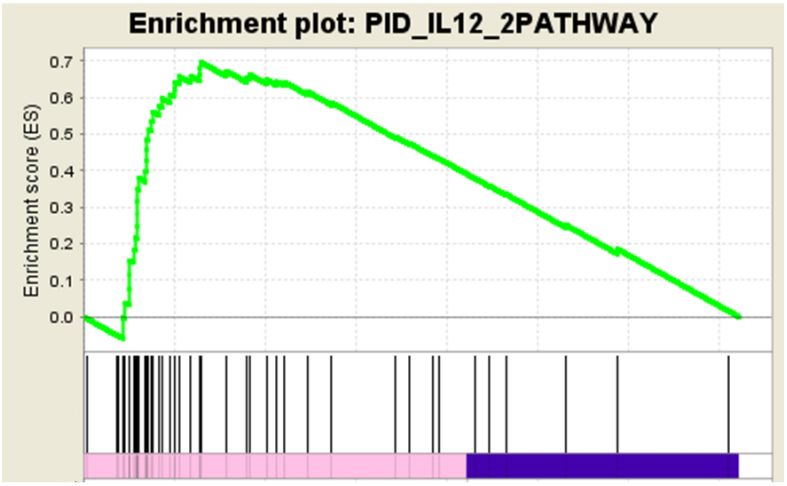


B)



Enrichment plot for canonical pathway:	FWER value	NES
Reactome: NCAM1 interactions	<0.000	-2.71
Reactome: Collagen formation	<0.000	-2.62
KEGG: Basal cell carcinoma	<0.000	-2.56
Reactome: ECM organization	<0.002	-2.45
KEGG: Dilated cardiomyopathy	<0.006	-2.35
KEGG: Hedgehog signaling pathway	<0.006	-2.34
PID: WNT signaling pathway	<0.017	-2.28
KEGG: Arrhythmogenic right ventricular cardiomyopathy ARVC	<0.018	-2.28
KEGG: ECM receptor interaction	<0.018	-2.28
PID: Integrin1 pathway	<0.020	-2.27
KEGG: Hypertrophic cardiomyopathy HCM	<0.029	-2.24
Reactome: muscle contraction	<0.029	-2.24
Reactome: NCAM signaling for neurite out growth	<0.044	-2.21

C)



Enrichment plot for canonical pathway:	FWER value	NES
KEGG: T cell receptor signaling pathway	<0.000	2.81
PID: IL12 pathway	<0.000	2.77

KEGG: Natural killer cell mediated cytotoxicity	<0.000	2.72
REACTOME: Innate immune system	<0.000	2.68
KEGG: Primary immunodeficiency	<0.000	2.68
REACTOME: Immunoregulatory interactions between a lymphoid and a non lymphoid cell	<0.000	2.65
KEGG: Chemokine signaling pathway	<0.000	2.63
KEGG: Cytokine-cytokine receptor interaction	<0.000	2.63
PID: TCR pathway	<0.000	2.62
KEGG: Toll-like receptor signaling pathway	<0.000	2.62
REACTOME: Immune system	<0.000	2.60
PID: CD8 TCR pathway	<0.000	2.59
KEGG: B cell receptor signaling pathway	<0.000	2.59

Figure 37 Gene set enrichment analysis of the sequencing dataset obtained from PDGFRα-positive cells RNA (P5 vs. W6). A) *Plots show additional information about the gene sets and enrichment results performed for canonical pathways (CP) and miRNA binding motifs (miR). Left histograms show the number of gene sets enriched at P5 (negative ES values) and W6 (positive ES values). Right panels show a global statistics of significant and non-significant gene sets. Plots of nominal p-values and FDR q-values (y-axis) vs. NES (x-axis) are shown.* ***B)*** *and* ***C)*** *Gene sets as a representation of canonical biological processes at P5 (negative ES values)* ***(B)*** *and W6 (positive ES values)* ***(C)****. Representative enrichment plots are shown for each time point. Green plots indicate an enrichment profile, where the peak indicates the ES for the gene set; black lines indicate the hits where the members of the gene set appear in the ranked list of genes (pink – genes encountered at W6; blue – genes encountered at P5). The top gene sets with the highest (W6) or lowest (P5) NES are shown in tables. NES= Normalized enrichment score, FWER= family-wise error rate, ES= enrichment score. References: [85]. Data was analyzed in collaboration with Dr. Mario Looso from Max Planck Institute in Bad Nauheim.*

Chapter 6

DISCUSSION

In the present study we characterized the miRNome and transcriptome of a mouse lung PDGFRα-positive cells. This fibroblast subset has been shown to be strongly involved in lung development, in particular in myofibroblast differentiation during septum formation [13, 14]. These results provide insights into new putative target molecules whose control might be a useful tool to manipulate lung regeneration or develop new therapeutic approaches for structural lung diseases.

6.1. Characterization of lung fibroblasts phenotype and localization in neonate and adult lung tissue

In the first part of this study, we investigated localization and phenotype of three fibroblast subsets, the lipofibroblasts, the myofibroblast and PDGFRα-positive cells. All three fibroblast subtypes play a crucial role during rodent lung alveolarization. In particular the myofibroblasts have been demonstrated to be involved in the process of secondary septation and thus the formation of new alveoli [81, 127]. We showed that the fibroblast subsets phenotypic markers' (ADRP, αSMA and PDGFRα) protein expression level decreases with the termination of alveolarization. Apoptosis of interstitial fibroblasts has been demonstrated to be one reason for this decrease in cell number [17, 102]. However the de-differentiation of fibroblasts subtypes might be a further reason [14]. Thus the regulation of the fibroblast phenotype plays a crucial role during lung development. Fibroblasts differentiation and proliferation need to be controlled in a tight and time-dependent manner. The myofibroblasts phenotypic marker, the α-smooth muscle actin (αSMA) protein and gene expression levels both peaked at postnatal day 10 as a consequence of "bulk alveolarization" process [19, 103]. However their expression was highly variable (at both protein and gene level) probably due to different developmental progress of the littermates used for cell isolation. The localization of myofibroblasts in the lung tissue was previously

well characterized. Their phenotypic markers, the αSMA or SMMHC are predominantly expressed at the tips of newly formed septa and around the alveolar rings during postnatal lung development [79]. However these two phenotypic markers are not exclusively specific for myofibroblast. They are also strongly expressed by smooth muscle cells surrounding bronchi/ bronchioles and blood vessels in neonate as well as in the adult lung. It has been shown that αSMA is regulated by different transcriptional mechanism in smooth muscle cells [39]. The slight changes of Pdgfrα gene expression levels at different lung developmental stages might suggest that its expression mainly is regulated on a post-transcriptional level. Post-transcriptional regulation of PDGFRα expression was already proposed in rodent epididymis studies [8]. PDGFRα was localized in the alveolar interstitium, around the bronchi/ bronchioles but clearly it was not expressed by the epithelium, the endothelium and leukocytes. This is in line with previous studies which showed that PDGFRα-positive cells are localized at the ridge of newly formed septa enabling differentiation into myofibroblasts during the process of alveolarization [79].

To get a clear picture of the fibroblast subsets phenotype, cells from whole lung homogenate were cultured at standard conditions and their differentiation capacity was investigated in detail. Fibroblasts obtained from neonatal animals had greater ability to differentiate into myofibroblasts than fibroblasts from adult mice. This observation was consistent with the fibroblast functionality in postnatal lung development. During this time, fibroblasts proliferate and differentiate to form new septa. These processes have to be tightly regulated by distinct molecular factors, which adjust the fibroblasts phenotype according to the requirements of the different lung developmental stages. In neonatal tissue fibroblasts are characterized by a “hyper-proliferative” phenotype whereas in adult healthy tissue they remain in a homeostatic “quiescent” phenotype. This in turn requires constant reprogramming of the fibroblasts phenotype and function which generally has been observed in cell culture [114]. Fibroblasts isolated from adult lung tissue differentiated into myofibroblasts, though they required more time for changing their phenotype than fibroblasts isolated from neonatal lung tissue. Therefore it is necessary to understand the mode of differentiation and the mechanisms leading to a change in phenotype of fibroblasts upon different requirements of physiological or pathophysiological processes. Against this

background, the molecular characterization of PDGFRα-positive fibroblasts is discusses in the following sections.

6.2. PDGFRα expression level in lung tissue and cell culture

The changes in phenotype and differentiation abilities of PDGFRα-positive cells were further investigated. The results showed that PDGFRα-positive cells (use of endogenous *Pdgfrα* gene tagged with GFP) are of mesenchymal origin and gain characteristics specific for myofibroblasts in cell culture. It has been shown that two different subsets of PDGFRα-positive cells are present in the lung, the dim and the bright [22, 55]. The "high-expressing" PDGFRα-GFP^{+} cells are associated with high expression of cytosolic αSMA, whereas "low-expressing" PDGFRα-GFP^{+} cells co-expressed αSMA on much lower level [55]. During lung development, the PDGFRα positive myofibroblasts are more resistant to apoptosis. [80]. Based on studies using the post-pneumonectomie re-alveolarization model, Chen et al. hypothesized that the PDGFRα-GFP^{dim} expression promotes the contractile function of fibroblasts, whereas the PDGFRα-GFP^{bright} expression promotes synthesis of structural proteins (synthetic phenotype). Further, they stated that during lung development and regeneration, a balance between the contractile and the synthetic PDGFRα-positive fibroblast phenotypes must be maintained [22]. Using immunocytochemistry of cultured lung fibroblasts, isolated from the same transgenic mouse line, we showed that the signals of PDGFRα and αSMA were less stable than the ADRP signal during the period of cell culture. When cultured shortly (2h or 24h), PDGFRα-GFP$^{bright+}$ cells were smaller and predominantly αSMA-negative but ADRP-positive, when cultured longer (72h), the cells were PDGFRα-GFP^{dim+} and double αSMA-, ADRP- positive. Moreover, using cells from wild type mice, we showed that PDGFRα$^{bright+}$ cells form long cytoplasmic extensions and may start to realize αSMA with high expression intensity what leads to new α-actin cytoskeleton formation. Thus, both αSMA and PDGFRα expression intensity decreased uniformly with myofibroblast cytoskeleton shaping. These data agree with the hypothesis of Chen et al.,; PDGFRα-GFP^{dim+} cells were characterized as differentiated homogenous myofibroblasts population, whereas PDGFRα-GFP$^{bright+}$ cells as undifferentiated fibroblasts with the ability to myofibroblasts

differentiation and ECM production. During lung development, the PDGFRα$^{bright+}$ cells represent the majority of PDGFRα-positive cells. On the other hand, in adult mice we were still able to identify a small number of quiescent PDGFRα$^{bright+}$ cells (FACSorting data). Thus, it may indicate that myofibroblasts differentiation and their cytoskeleton formation is strongly dependent on PDGFRα expression level regulation during lung development and in adulthood.

6.3. MiRNA profiles of pulmonary PDGFRα-positive cells across different time points of lung development

We aimed to understand the regulatory mechanisms which orchestrate the function and the phenotype of PDGFRα-positive fibroblast during lung septation and later in the adult tissue. To select the PDGFRα-positive cells from whole lung cells, we used magnetic beads/anti- PDGFRα antibody selection, thus only the cells with extracellularly exposed N-terminal domain of the receptor were sorted. We chose four specific time points of lung development: E16.5, P5, P14 and W6. At E16.5 the process of alveolarization has not started yet while at W6 it is completed. In contrast, at P5 the alveolarization has just started, and at P14 alveoli are already formed and the process of microvascular maturation and thinning of septal interstitium occurs. Therefore, the postnatal day 5 was the most interesting time point to us, as at this time the major changes necessary for gas exchange units (alveoli) formation begin [19, 56, 92, 106]. Subsequently, we analyzed the miRNome of PDGFRα-positive fibroblast. It is known that miRNAs are the key regulators involved in many biological processes, such as apoptosis, differentiation and proliferation [26]. Using TaqMan low density arrays (TLDAs), we identified 98 miRNAs which were differentially expressed across the four time points. Moreover, these miRNAs could be split into specific clusters based on their similar expression patterns within the four time points. 56 of these miRNAs were up-regulated at P5 vs. W6 and only 16 miRNAs were up-regulated at W6 vs. P5. Previously it has been shown, that miRNAs may exhibit temporal and spatial cell type specific expression patterns and many of them are strongly involved and broadly regulated during lung development [11, 31] or lung diseases [123]. Indeed in our data miRNAs up-regulated at postnatal day 5 were associated with developmental processes while at week 6

they were associated with the emergence of respiratory and connective tissue diseases or inflammatory response. Interestingly, there was no substantial difference between those two miRNA groups when it comes to regulation of the basic cell processes. It shows that during lung development and in adulthood, different miRNAs may act to confer accuracy and uniformity of the PDGFRα-positive cells phenotype and functions [35].

To identify target genes of the miRNAs, we used some of the available online algorithm software which have been developed over the past few years [7]. This approach gave us helpful hints for further experiments. Using miRTarBase [50], we searched for the miRNAs validated to target the Pdgfrα gene. In our study none of these miRNAs was differentially regulated in PDGFRα-positive cells across the four time points of lung development. Nevertheless, we found 4 miRNAs up-regulated at P5 (miR-421, -449b, -149, -20b) which were predicted to target *Pdgfrα* 3'-UTR within its highly conserved sequence sites among vertebrates. Further miRNA bio-function analysis suggested that these 4 miRNAs might be strongly implicated in regulation of fibroblast differentiation during alveolarization. However more information about these miRNAs is required. *In vitro* or *in vivo* functional studies would be the next step to investigate a gain or loss (use of mimics or inhibitors) of these miRNAs functions what simultaneously could prove their potential therapeutic application in lung interstitial diseases.

Using miRTarBase [50], we searched for other miRNA candidates regulated in PDGFRα-positive cells, which are validated to target the genes that are involved in fibroblast subsets differentiation or ECM formation. We identified 2 miRNAs that are validated to target *Tgfbr2* (miR-370-3p, -204-5p; both up-regulated at P5 vs. W6) [73, 120], 1 validated to target *Tgfb1* and *2* (miR-29b-3p; constant at P5 and W6) [75], 1 validated to target *Pparγ* (miR-20b-5p; also predicted to target *Pdgfrα*; up-regulated at P5 vs. W6) [45], 2 validated to target elastin (miR-29a: down-regulated at P5 vs. W6 and miR-29b: constant at P5 vs. W6) [89, 118], and 3 validated to target different types of collagen (miR-29a, b, c; miR-29c: down-regulated at P5 vs. W6) [33, 44, 69, 118]. In fact, expression kinetics of all these miRNAs was opposite to the expression level of their target genes which analysis was performed based on our sequencing data. Hence, it is possible, that these miRNAs repress the expression of these genes. Further functional studies of these molecules (knock-down or overexpression) is needed to understand how these relationships are implicated in the

regulation of the process of alveolarization and the quiescent phenotype of fibroblasts in the adult lung.

6.4. Differentially expressed genes of pulmonary PDGFRα-positive cells of postnatal and adult time points

Next-generation sequencing was used to analyze the transcriptome of lung PDGFRα-positive cells and negative control obtained from postnatal day 5 and week 6 mice. Using gene set information and canonical pathway analysis we identified gene candidates differentially regulated at P5 and W6. At P5 most of the genes were related with developmental processes and cellular function, whereas at W6 they were related with immunological response, cytokines signaling and transcription/translation regulation. Similarly, analysis of differentially regulated genes between P5 vs. W6, P5 vs. NC and W6 vs. NC showed that all the genes responsible for myofibroblasts differentiation and the synthetic phenotype (ECM components production), such as *Acta2*, *Myh11*, types of *Myl*, *Pdgfrα*, *Pdgfrl*, types of *Col*, *Eln*, *Ctgf*, *Tgfbi*, *Tgfb2*, *3*, types of *Ltbp* and *Tgfbr3* were down-regulated in the PDGFRα-positive cells obtained at W6 vs. P5. Most of these genes were previously reported and closely investigated. It has been shown that TGFβ induces differentiation of lung fibroblasts into myofibroblasts and it is implicated in the pathogenesis of pulmonary fibrosis [62]. Another gene, *Tgfbi* was highly up-regulated at both P5 and W6 comparing to NC, and it was log2 (1.18)-fold change higher at P5 vs. W6. The *Tgfbi* gene encodes the extracellular matrix (ECM) attachment protein which contains an RGD (DNA recognition sequence) site. It binds to the type I, II and IV collagens (cell-collagen interactions) and serves as a ligand recognition sequence for several integrins [87]. Thus it seems to be necessary for the cell-matrix interactions, cell adhesion, migration and differentiation during the process of alveolarization. Additionally, two miRNA genes i.e. *mir-487b* and *mir-382* were up-regulated at P5 vs. NC and significantly down-regulated at W6 vs. P5 according to our data obtained by the use of the TLDA method. Validated gene targets and functions of these two miRNAs are still not well characterized. Recent studies showed that cigarette smoke mediates epigenetic repression of miR-487b during human pulmonary carcinogenesis (study based on epithelial cell culture) and revealed its 5 direct

targets: SUZ12, BMI1, WNT5A, MYC, and KRAS [125]. Whereas miR-382 contributes to renal inner medullary interstitial fibrosis in a mouse model and it targets kallikrein 5 (*Klk5*) [60]. Thus, these two miRNAs could be an interesting indication for further functional studies of the lung during development or disease models.

Our data suggest that the gene profile of PDGFRα-positive cells isolated from W6 mouse lungs is more characteristic for the lipofibroblast phenotype rather than the myogenic phenotype. Cells isolated from W6 lung tissue were characterized by the higher expression of following genes: *Pparγ*, *Rxra*, *Thy1*, *Lpl* (all four genes were previously reported as lipofibroblasts markers), *Cotl1* and *mir-29a*. It has been shown that THY-1 signals through PPARγ/RXRa to promote lipofibroblast differentiation during lung development [77, 78, 108, 119] and antagonizes the fibrogenic processes, like cell differentiation, inflammation and wound healing as well as myofibroblast differentiation [18, 64, 98, 119]. On the other hand, we showed that fibrogenic *Tgfβr1* and *2* were also up-regulated at W6 vs. P5. However, their activation might be inhibited by PPARγ signaling via a Smad-independent pathway [62]. Substantial up-regulation of *mir-29a* at W6 vs. P5 was also consistent with our data obtained by the use of TLDA. Mir-29a might have an crucial effect on elastin [89, 118] and collagen [33] negative regulation, whereas coactosin-like protein (*Cotl1*; actin-binding protein) might negatively regulate F-actin cytoskeleton formation in myofibroblasts [32, 93].

Previous multiple studies showed that the pulmonary lipofibroblasts are THY1+ PPARγ+ and number of these cells increases at the onset of alveolarization [3], however not all of the lipofibroblasts are $THY1^+$ [119]. In contrast, our data showed that *Thy1*, *Rxra* and *Pparγ* genes expression level was up-regulated after lung development termination, i.e. in adulthood rather than in neonate. *Thy1* gene expression pattern from our data was consistent with previous observations, showing that in rodent lung tissue homogenate the $THY1^+$ cell number increases at P5 and then even more in adult [119]. In spite of this, $THY1^+$ cells might not play a critical role in the surfactant production. It has been shown that *Thy1* knockout mouse pups do not exhibit surfactant deficiency or respiratory distress syndrome [119]. Furthermore, *Adrp* was not found to be differentially expressed between P5, W6 and even NC. Previously it has been shown that ADRP is essential for many metabolic processes of different cell types, like macrophages, cultured epithelial cells or

within tissue by lactating mammary epithelial cells [46, 95]. Therefore, the ADRP might not be a perfect marker for lipofibroblasts like thought before [94, 104]. In our studies, the ADRP protein expression level was significantly up-regulated (~2.5-fold change) in whole lung tissue homogenate at P5 vs. adult (W8). Its expression level might be decreased due to the apoptosis of fibroblasts after the alveolarization process termination. On the other hand, the solute carrier family 27A6 (*Slc27a6*; membrane transport protein) responsible for fatty acid transport was log2 (4.06)-fold change higher at P5 vs. W6. This finding supports the concept that fatty-acid transporter proteins might mediate increased fatty-acid uptake necessary for surfactant production during alveolarization [119]. Furthermore, our data suggest that THY1 and PPARγ signaling pathway may have a crucial role in PDGFRα-positive cells differentiation regulation in adulthood. Increased *Thy1* and *Pparγ* gene expression might negatively influence myofibroblast phenotype induction by delaying PDGFRα activation or antagonizing *Pdgfrα* gene expression [22].

Our data showed that *Pdgfrα* gene expression level was log2 (2.7) down-regulated at W6 vs. P5 and log2 (2.2) down-regulated at W6 vs. NC. It was not 2-fold change different at P5 vs. NC. This fact was striking since PDGFRα-positive cells were selected via anti-PDGFRα antibody positive selection and negative control cells via anti-PDGFRα antibody negative selection. One explanation might be that in the negative control the present cell types (presumably it could be already existing number of myofibroblasts or smooth muscle cells) did not express PDGFRα at the cell surface, but significant *Pdgfrα* gene expression might have led to intracellular protein expression which was post-transcriptionally regulated or the PDGFRα protein was recycled or antagonized. These findings are supported by the fact, that numerous types of negative regulatory mechanism of RTKs have been already proposed, i.e. the ligand and receptor availability, regulation of downstream signaling components or of the respective transcription factors, post-transcriptional regulation or the receptor endocytosis (receptor recycling or degradation) [20]. Abundance of unstimulated tyrosine kinase receptors (RTK) at the cell surface may be controlled via production and continuous recycling. Interactions between cytoskeletal components and the plasma membrane may also affect receptor distribution and signaling by forming cytosolic corral traps (compartmentalization of receptor within the plasma membrane). Additionally, it has been shown that RTK signaling is not only limited to the

cell surface, but can continue within the endocytic compartment [20]. As mentioned before, PDGFRα is the receptor responsible for conducting signals necessary for basic biological function of the cell, ranging from cell migration and survival to proliferation and differentiation. Therefore, it must be under tight spatial regulation enabling tissue morphogenesis and function during lung development. However it remains unsolved how exactly the receptor signaling is regulated. We already showed that 4 miRNAs (miR-421, -449b, -149, -20b) which are highly expressed by PDGFRα-positive cells at P5 might putatively target the *Pdgfrα* 3'-UTR within its highly conserved sites. This could be one of the possible post-transcriptional regulatory mechanisms of *Pdgfrα* gene expression during development. Furthermore, previously a large number of inhibitory proteins that act to attenuate the signals originating from activated RTK receptors have been identified. These specific proteins are working as negative feedback loop mechanisms and can be present prior to the receptor activation or are transcriptionally induced upon the receptor stimulation in the cell. One of these proteins called Sprouty (SPRY) is transcriptionally activated and later phosphorylated by RTK signaling. Sproutys are known inhibitors of fibroblast growth factor (FGF) signaling and play an essential role in lung branching morphogenesis [43]. However, only SPRY4 (but not SPRY1 and SPRY2) phosphorylation can be induced by FGF, EGF, and PDGF [111]. Our data showed that two gene members of the *Spry* family were up-regulated at P5 comparing to W6 and down-regulated at W6 comparing to NC. The *Spry4* expression level was log2 (2.34)-fold change and *Spry1* log2 (2.65)-fold change higher at P5 vs. W6, while *Spry4* was log2 (2.90)-fold change and *Spry1* log2 (2.39)-fold change lower at W6 vs. NC. These findings show that the *Spry* genes expression level was elevated consistently with *Pdgfrα* gene expression level in PDGFRα-positive cells isolated from P5 mice and in NC. In turn, this suggests the involvement of a negative feedback loop regulatory mechanism of the *Pdgfrα* gene at P5 but not W6. Likely, the low expression level of the *Pdgfrα* gene at W6 might be an effect of PPARγ-dependent regulation and in consequence the lack of PDGFRα protein activation might induce a dispensable function of the *Spry* genes. Higher expression level of *Pparγ* gene at W6 could also have an effect on the activation of *Spry4* expression. In the same line Tennis, et al. showed that *Spry4* expression level is regulated by PPARγ signaling outside of the RTK pathway [111]. Analysis of PDGFRα-positive cells gene and miRNA expression profiles

allowed us to identify putative molecular mechanisms involved in PDGFRα-positive fibroblasts regulation and myofibroblasts differentiation that vary in neonate and adulthood.

Chapter 7

CONCLUSIONS

Pulmonary PDGFRα-positive fibroblasts have been shown to obtain a high plasticity and capacity to differentiation into specialized fibroblasts subtypes during the process of alveolar septation. But the underlying regulatory mechanisms still remain unknown. Furthermore a variety of molecular switches might be involved in the signaling during alveolar septation. Since pulmonary structural diseases are characterized by a loss or destruction of the alveolar structures, a better understanding of these molecular mechanisms driving alveolar septum formation might help to discover novel targets for therapeutic regenerative approaches. Therefore the present study aimed to compare the molecular profile at the miRNome and transcriptome level of PDGFRα-positive fibroblasts at different time points of lung development

The identification of lung PDGFRα-positive fibroblast specific miRNA expression profiles revealed a distinct number of relevant candidates. In particular, 98 miRNAs were identified and characterized by different expression patterns across four time points of lung development. 56 of these miRNAs were up-regulated at P5 vs. W6 and only 16 miRNAs at W6 vs. P5, whereas the remaining miRNAs were expressed at the same level at P5 and W6. *In silico* analysis of these miRNAs revealed more insight into PDGFRα-positive fibroblasts function and phenotype during alveolarization. Next-generation sequencing supported our data from the miRNA analysis and disclosed further relevant gene candidates. Transcriptome data analysis showed that PDGFRα-positive cells molecular phenotype and differentiation abilities highly differ during the onset of alveolarization, and later in adulthood.

In conclusion, we discovered relevant miRNA and gene candidates which represent the first basis for further loss- or gain-of-function studies to identify new therapeutic targets for structural lung diseases.

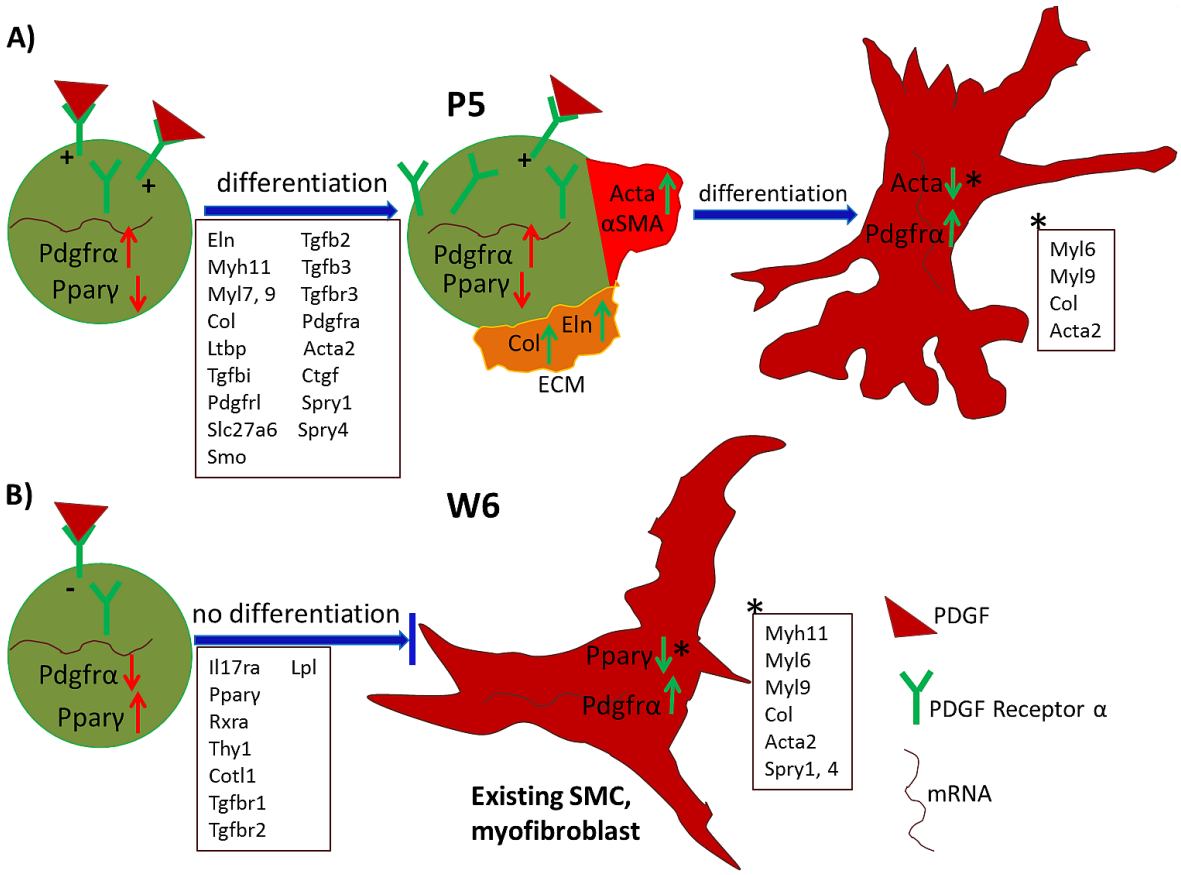


Figure 38 Schematic picture of PDGFRα-positive fibroblasts phenotype and their differentiation capacity at postnatal day 5 (P5) and week 6 (W6). *Arrows indicate the level of differentially expressed genes (up↑, down↓). In the boxes, genes differentially up-regulated at a specific stage of cell differentiation or time point are shown. "+" indicates activated PDGF Receptor α signaling, "-" indicates lack of the activation.* ***A)*** *At P5, the PDGFRα-positive fibroblasts differentiate into myofibroblasts and produce ECM (Col= collagen, Eln= elastin). Hypothetical phenotype of differentiated myofibroblast is shown (*).* ***B)*** *At W6 the PDGFRα-positive fibroblasts do not differentiate into myofibroblast and remain in quiescent mode. Myofibroblasts and smooth muscle cells (SMCs) differentiated during late lung development are still present at W6 ("existing SMC, myofibroblast"). Hypothetical phenotype of myofibroblast and SMC is shown (*).*

ABSTRACT

BACKGROUND AND AIM The process of alveolarization is tightly regulated and requires the contribution of different subpopulations of fibroblasts such as myofibroblasts, lipofibroblasts and platelet-derived growth factor receptor α (PDGFRα)-positive fibroblasts. Each of this fibroblasts subset fulfills certain functions during lung development in a time-dependent manner. In particular PDGFRα-positive cells are crucial for alveolar septation and myofibroblasts differentiation. PDGF-A deletion *in vivo* leads to a disruption of septum formation in mice. Therefore, we aimed to characterize PDGFRα-positive fibroblasts specific miRNA and mRNA expression profiles at different time points of lung development. We hypothesized that their phenotype, differentiation capacities and functions might be regulated by the change of expression profiles during the different stages of lung development and homeostasis.

RESULTS Lung PDGFRα-positive fibroblasts were isolated using magnetic beads/anti-PDGFRα antibody positive selection. Using TaqMan low density array analysis we identified 98 significantly regulated miRNAs out of 750 tested miRNAs that exhibited different expression patterns across four time points of lung development. 56 of these miRNAs were up-regulated at P5 vs. W6 and only 16 miRNAs at W6 vs. P5. *In silico* analysis of these miRNA candidates allowed us to identify their function and gene targets. Subsequent analysis of the PDGFRα-positive fibroblasts transcriptome using next-generation sequencing revealed gene candidates differentially regulated during alveolarization and in adulthood.

CONCLUSIONS MiRNome and transcriptome dataset analysis showed that the molecular phenotype and functions of pulmonary PDGFRα-positive cells greatly differ during the onset of alveolarization, and later in adulthood. Simultaneously, we identified relevant miRNAs and gene candidates which might serve as a basis for “loss or gain of function” analysis. These functional studies are essential to identify new therapeutic targets for structural lung diseases and regenerative medicine.

ZUSAMMENFASSUNG

HINTERGRUNDINFORMATION und ZIEL Der Prozess der Alveolarisierung unterliegt einer strengen Regulation. Verschiedene Subpopulationen pulmonaler Fibroblasten wie beispielsweise Myofibroblasten, Lipofibroblasten und platelet-derived growth factor receptor α (PDGFRα)-positive Fibroblasten sind daran beteiligt. Jeder dieser Fibroblasten Subtypen vollbringt eine zeitabhängige bestimmte Funktion während der Lungenentwicklung. Insbesondere tragen PDGFRα-positive Zellen wesentlich zu der Septenbildung und der Differenzierung von Myofibroblasten bei. *In vivo* Deletion von PDGF-A führt zu einer Störung der Alveolarisierung in Mäusen. Daher liegt das Ziel der vorliegenden Arbeit darin, miRNA- und mRNA-Expressionsprofile PDGFRα-positiver Fibroblasten, zu verschiedenen Zeitpunkten der Lungenentwicklung zu bestimmen. Die Hypothese der vorliegenden Arbeit besteht darin, dass der Phänotyp, die Differenzierungsmöglichkeiten und die Funktionen PDGFRα-positiver Fibroblasten durch eine Änderung der Expressionsprofile zu verschiedenen Zeitpunkten der Lungenentwicklung und Homöostase reguliert werden.

ERGEBNISSE PDGFRα positive Fibroblasten der Lunge wurden durch Magnetseparation und positiv Selektion mittels Antikörper gegen PDGFRα isoliert. Die miRNome-Analyse der PDGFRα-positiven Fibroblasten an vier verschiedenen Zeitpunkten der Lungenentwicklung durch TaqMan low density array ergab 98 statistisch signifikant regulierte miRNAs von 750 analysierten miRNAs. 56 dieser miRNAs zeigten sich hoch-reguliert an P5 vs. W6 und nur 16 miRNAs an W6 vs. P5. *In silico* Analysen dieser miRNA Kandidaten ermöglichte es, ihre Funktionen und Zielgene zu identifizieren. Die Analyse des Transcriptoms PDGFRα-positiver Fibroblasten durch “next-generation sequencing” erbrachte Gen Kandidaten deren Expression während der Alveolarisierung im Vergleich zu der adulten Lunge unterschiedlich reguliert war.

SCHLUSSFOLGERUNG Die miRNome und Transcriptome Datenanalyse zeigte, dass sich der molekulare Phänotyp und die Funktionen von PDGFRα-positiven Zellen zwischen dem Beginn der Alveolarisierung und später der adulten Lunge sehr stark unterscheiden. Gleichzeitig wurden relevante miRNAs und Gen Kandidaten identifiziert, die als Ansatzpunkt für „loss“ oder „gain of function” Analysen dienen können. Diese funktionellen Studien sind erforderlich, um therapeutisch relevante Zielmoleküle für die regenerative Therapie von Lungengerüsterkrankungen zu identifizieren.

ABBREVIATIONS

°C – Celsius degree

µg – microgram

µl – microliter

µm – micrometer

AGCT – adenine, guanine, cytosine, thymine

Asp – aspartic acid

bp – base pair

BSA – bovine serum albumin

CCD – charge-coupled device

CD – cluster differentiation

cDNA – complementary deoxynucleic acid

cm – centimeter

CO_2 – carbon dioxide

CreER – tamoxifen-dependent (binds to estrogen receptor (ER)) Cre recombinase

Cy3 – cyanine 3

DAPI – 4',6-diamidino-2-phenylindole

ddATP – 2',3'-Dideoxyadenosine-5'-Triphosphate

ddCTP – 2',3'-Dideoxycytidine-5'-Triphosphate

ddGTP – 2',3'-Dideoxyguanosine-5'-Triphosphate

ddH_2O – double distilled water

ddTTP – 2',3'-Dideoxythymidine-5'-Triphosphate

DNA – deoxynucleic acid

DNase – deoxynuclease

EDTA – ethylenediaminetetraacetic acid

eGFP – enhance green fluorescent protein

FITC – fluorescein isothiocyanate

Gb – giga base pairs

gDNA – genomic deoxynucleic acid

h – hour

H2B – histone H2B

HCl – hydrochloric acid

His – histidine

IgG – immunoglobulin G

kDa – kilodalton

KH_2PO_4 – potassium dihydrogen phosphate

$KHCO_3$ – potassium bicarbonate

loxP – locus of X-over P1 is a site on the Bacteriophage P1 consisting of 34 bp

M – molar

min – minute

ml – milliliter

mM – milimolar

mRNA – messenger ribonucleic acid

$Na_2HPO_4 \cdot 2H_2O$ – disodium phosphate dihydrate

NaCl – sodium Chloride

NH_4Cl – ammonium chloride

nt – nucleotide

PBS – phosphate buffered saline

PCR – polymerase chain reaction

PFA – paraformaldehyde

pH – hydrogen ion concentration

qPCR – quantitative polymerase chain reaction

RNA – ribonucleic acid

rpm – revolutions per minute

rRNA – ribosomal ribonucleic acid

RT – room temperature

SD – standard deviation

SDS – sodium dodecyl sulfate

SDS-PAGE – sodium dodecyl sulfate polyacrylamide gel electrophoresis

TAE – tris-acetate-EDTA

TaqMan – Taq Polymerase + PacMan = TaqMan

TBST – tris-buffered saline and tween 20

TEMED – tetramethylethylenediamine

UNG – uracil N-glycosylase

UV – ultraviolet

V – volt

LIST OF FIGURES

LIST OF TABLES

APPENDIX

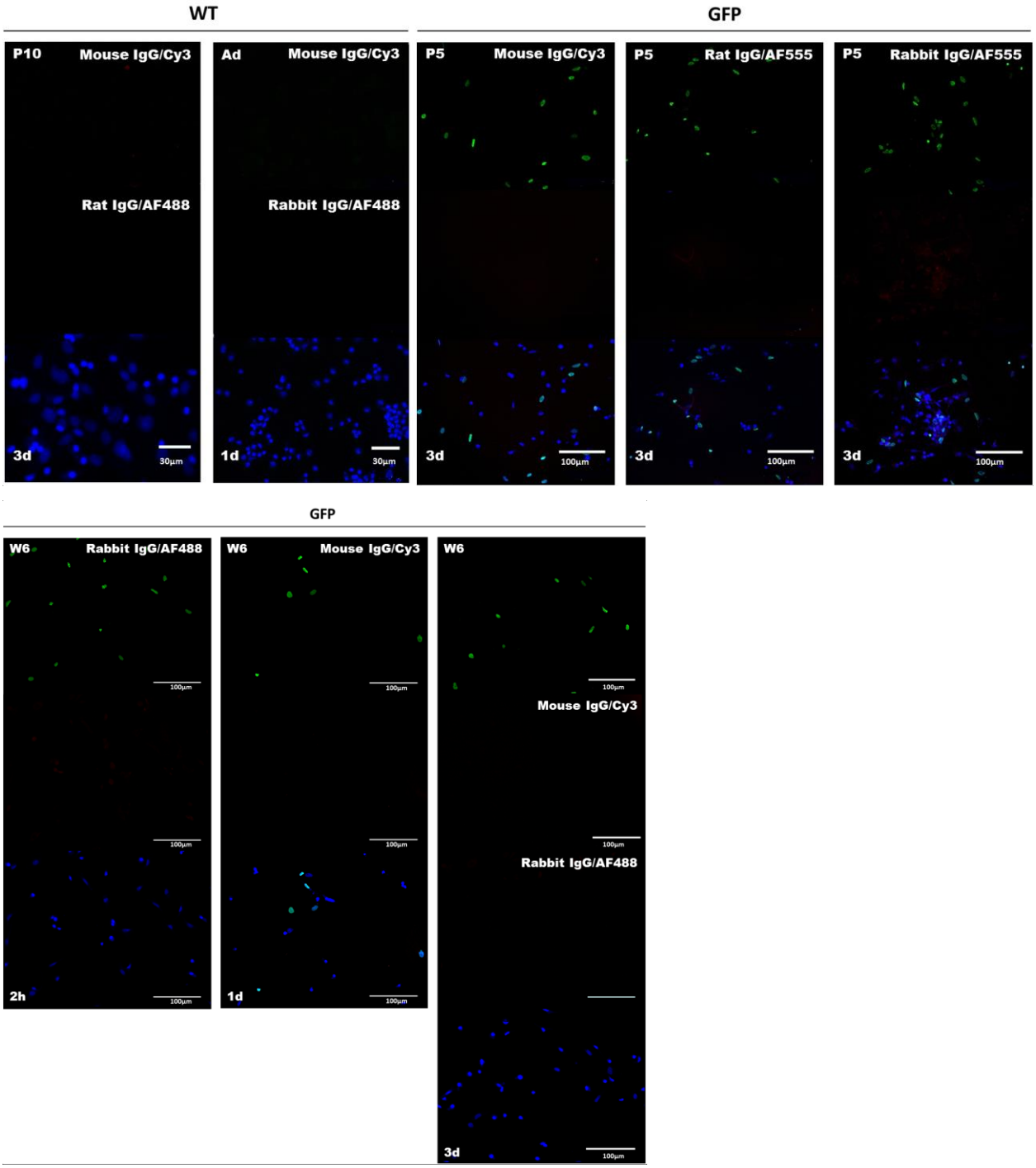


Figure A1 Isotype control of immunofluorescent staining in cell culture. *Wild type (WT) or GFP-tagged mouse lung cells were stained with mouse IgG/Cy3, rat IgG or rabbit IgG primary antibodies and Alexa Fluor 488 or 555 secondary antibodies. Scale bar = 30 or 100μm.*

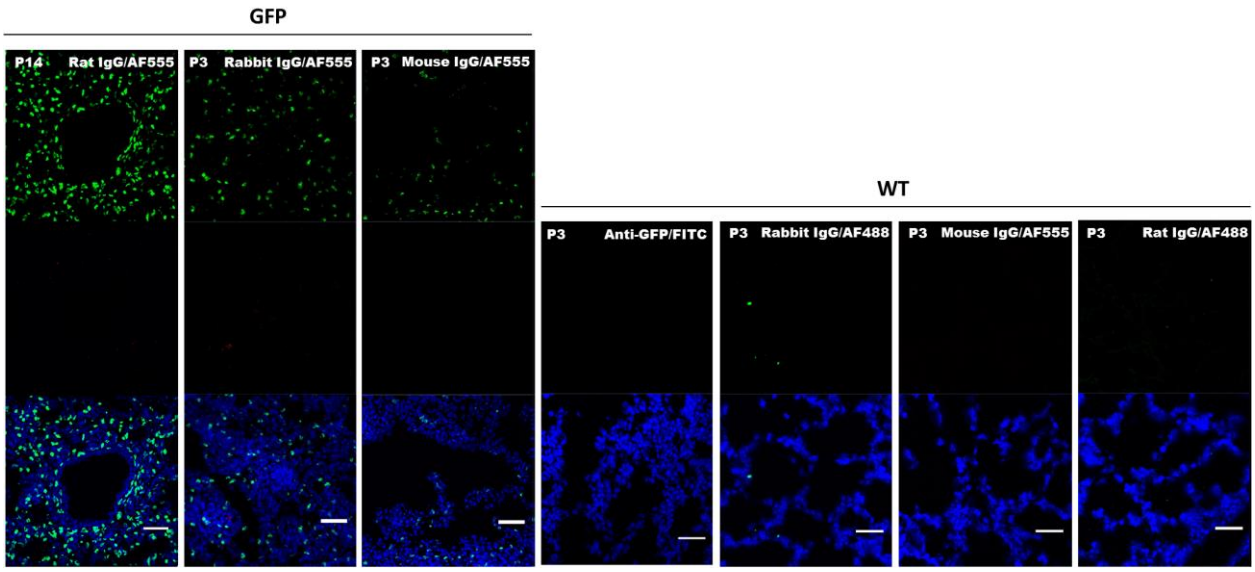


Figure A2 Isotype control of immunofluorescent staining in lung tissue. *Wild type (WT) or GFP-tagged mice lung tissue slides were stained with anti-GFP/FITC, mouse IgG, rat IgG or rabbit IgG primary antibodies and Alexa Fluor 488 or 555 secondary antibodies. Scale bar = 50µm.*

Table A1 Affiliation of miRNAs into miRNA families based on the same seed region sequence. According to IPA.

CLUSTER	miRNAs	miRNA family w/in seed	CLUSTER	miRNAs	miRNA family w/in seed
Cluster 0	mmu-miR-666-5p	miR-666-5p w/seed GCGGGCA	**Cluster15**	mmu-miR-370	miR-370-3p w/seed CCUGCUG
				mmu-miR-433	miR-433-3p w/seed UCAUGAU
				mmu-miR-434-5p	miR-434-5p w/seed CUCGACU
				mmu-miR-495	miR-495-3p w/seed AACAAAC
				mmu-miR-337	miR-337-3p w/seed UCAGCUC
				mmu-miR-487b	miR-487b-3p w/seed AUCGUAC
				mmu-miR-541	miR-541-5p w/seed AGGGAUU
				mmu-miR-543	miR-543-3p w/seed AACAUUC
				rno-miR-350	miR-350 w/seed UCACAAA
				rno-miR-551B	miR-551b-3p w/seed GCGACCC
Cluster2	mmu-miR-543	miR-543-3p w/seed AACAUUC	**Cluster16**	mmu-miR-30d	miR-30c-5p w/seed GUAAACA
	mmu-miR-299	N/A		mmu-miR-146a	miR-146a-5p w/seed GAGAACU
	mmu-miR-376a*	miR-376a-5p w/seed GUAGAUU)		mmu-miR-146b	miR-146a-5p w/seed GAGAACU
	mmu-miR-673-3p	miR-673-3p w/seed CCGGGGC		mmu-miR-150	miR-150-5p w/seed CUCCCAA
				mmu-miR-222	miR-221-3p w/seed GCUACAU
				mmu-miR-29a	miR-29b-3p w/seed AGCACCA
				rno-miR-146B	miR-146a-5p w/seed GAGAACU
Cluster 4	hsa-miR-421	miR-421-3p w/seed UCAACAG	**Cluster 17**	mmu-miR-369-5p	miR-369-5p w/seed GAUCGAC
	mmu-miR-1939	N/A			
	mmu-miR-674*	miR-674-3p w/seed ACAGCUC			
Cluster5	mmu-miR-323-3p	mir-323-3p w/seed ACAUUAC	**Cluster18**	mmu-miR-126-3p	miR-126-3p w/seed CGUACCG
	mmu-miR-410	miR-344d-3p w/seed AUAUAAC		mmu-miR-223	miR-223-3p w/seed GUCAGUU
	mmu-miR-130b*	miR-130b-5p w/seed CUCUUUC		hsa-miR-200c	miR-200b-3p w/seed AAUACUG
	mmu-miR-376b*	miR-376b-5p w/seed UCAUAGA		hsa-miR-223	miR-223-3p w/seed GUCAGUU
Cluster6	mmu-miR-805	N/A	**Cluster19**	mmu-miR-127	miR-127-3p w/seed CGGAUCC
				mmu-miR-193b	miR-193a-3p w/seed ACUGGCC
				mmu-miR-296-5p	miR-296-5p w/seed GGGCCCC
				mmu-miR-335-3p	miR-335-3p w/seed UUUUCAU
				mmu-miR-335-5p	miR-335-5p w/seed CAAGAGC
				rno-miR-351	miR-125b-5p w/seed CCCUGAG
				mmu-miR-411	N/A
				hsa-miR-149	miR-149-5p w/seed CUGGCUC

				hsa-miR-93* mmu-miR-1839-3p mmu-miR-322* mmu-miR-322 mmu-miR-503* rno-miR-20b	miR-93-3p w/seed CUGCUGA miR-1839-3p w/seed GACCUAC miR-322-3p w/seed AACAUGA N/A miR-503-3p w/seed AGUAUUG miR-17-5p w/seed AAAGUGC
Cluster 7	mmu-miR-29c mmu-miR-139-5p mmu-miR-155 mmu-miR-34c* rno-miR-664	miR-29b-3p w/seed AGCACCA miR-139-5p w/seed CUACAGU miR-155-5p w/seed UAAUGCU miR-34c-3p w/seed AUCACUA miR-664-3p w/seed AUUCAUU	**Cluster20**	mmu-miR-126-5p mmu-miR-92a hsa-miR-200b mmu-miR-374-5p	miR-126-5p w/seed AUUAUUA miR-92a-3p w/seed AUUGCAC miR-200b-3p w/seed AAUACUG miR-374b-5p w/seed UAUAAUA
Cluster 8	mmu-miR-451	miR-451a w/seed AACCGUU	**Cluster 22**	mmu-miR-204 mmu-miR-376a mmu-miR-376b hsa-miR-206	miR-204-5p w/seed UCCCUUU miR-376a-3p w/seed UCGUAGA miR-376a-3p w/seed UCAUAGA miR-1w/seed GGAAUGU
Cluster 9	mmu-miR-1904	miR-1904 w/seed UUCUGCU	**Cluster24**	mmu-miR-463	N/A
Cluster 10	mmu-miR-135b mmu-miR-376c mmu-miR-379 mmu-miR-382 mmu-miR-409-3p mmu-miR-434-3p hsa-miR-214* hsa-miR-455 mmu-miR-449b rno-miR-409-3P Y1	miR-135a-5p w/seed AUGGCUU miR-376c-3p w/seed ACAUAGA miR-379-5p w/seed GGUAGAC miR-382-5p w/seed AAGUUGU miR-409-3p w/seed AAUGUUG miR-434-3p w/seed UUGAACC mir-214-5p w/seed GCCUGUC miR-455-5p w/seed AUGUGCC mir-34a-5p w/seed GGCAGUG miR-409a-3p w/seed AUGUUGC N/A	**Cluster25**	mmu-miR-29b mmu-miR-23a mmu-miR-1931 mmu-miR-1943	miR-29b-3p w/seed AGCACCA miR-23a-3p w/seed UCACAUU miR-1931 w/seed UGCAAGG miR-1321 w/seed AGGGAGG
Cluster13	mmu-miR-375 mmu-miR-532-3p	miR-375-3p w/seed UUGUUC miR-532-3p w/seed CUCCCAC	**Cluster27**	mmu-miR-205 mmu-miR-423-5p	miR-205-5p w/seed CCUUCAU miR-423-5p w/seed GAGGGGC
Cluster14	mmu-miR-329 rno-miR-758 rno-miR-219-1-3p hsa-miR-154* mmu-miR-1199 mmu-miR-483* mmu-miR-483 mmu-miR-879*	miR-329-3p w/seed ACACACC miR-758-3p w/seed UUGUGAC miR-219a-1-3p w/seed GAGUUGC miR-154-3p w/seed AUCAUACN/A miR-1199-5p w/seed CUGAGUCN/A miR-483-3p w/seed CACUCCU N/A miR-879-3p w/seed CUUAUGG	**Cluster 28**	hsa-miR-28-3p mmu-miR-31*	miR-28-3p w/seed ACUAGAU miR-31-3p w/seed GCUAUGC
			Cluster 29	mmu-miR-668 hsa-miR-136 rno-miR-379* rno-miR-382*	miR-668-3p w/seed GUCACUC miR-136-5p w/seed CUCCAUU miR-379-3p w/seed UAUGUAA N/A

Table A2 Nomenclature of genes or encoded proteins inside the miRNA networks generated by IPA.

Dicer1	Dcr-1 homolog	Tp53	Tumor protein p53
Ssb	Sjögren's syndrome antigen B (autoantigen La)	E2f3 E2f2 E2f1	E2F transcription factor 3 E2F transcription factor 2 E2F transcription factor 1
Dcp2	mRNA-decapping enzyme 2	Ventx	VENT homeobox
Ago2	Argonaute RISC catalytic component 2	Clpb	Caseinolytic peptidase B protein homolog
Nr0b2	nuclear receptor subfamily 0, group B, member 2	Ctnnb1	β-catenin (cadherin-associated protein)
Esrrg	Estrogen-related receptor gamma	Crebl2	(CRE)-binding protein-like-2
Ddx20	DEAD (Asp-Glu-Ala-Asp) box polypeptide 20	Inppl1	Inositol polyphosphate phosphatase-like 1
Zic1	Zic Family Member 1	Dusp9	Dual specificity phosphatase 9

Cdkn1b	Cyclin-dependent kinase inhibitor 1B (p27, Kip1)	Gprc6a	G protein-coupled receptor, family C, group 6, member A
Cdk7 Cdk5 Cdk4	Cyclin-dependent kinase 7 Cyclin-dependent kinase 5 Cyclin-dependent kinase 4	Smad2/3 Smad7	SMAD family member 2 and 3 complex SMAD family member 7
Map2k1/2 Map4k4 Mapk14	Dual specificity mitogen-activated protein kinase kinase 1/2 Mitogen-activated protein kinase kinase kinase kinase 4 Mitogen-activated protein kinase 14	Lin28A	Lin-28 homolog A
Pik3ca	Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha	Sgca	Sarcoglycan, alpha (50kDa dystrophin-associated glycoprotein)
Tgfb1	Transforming growth factor beta 1	Rps15	Ribosomal protein S15
Aqp4	Aquaporin 4	Ins1	Insulin I
Fhod1	Formin homology 2 domain containing 1	Ccna2 Ccne1 Ccnd1 Ccnd2	Cyclin A2 Cyclin E1 Cyclin D1 Cyclin D2
Erk	Extracellular signal-regulated kinases	Rb1	Retinoblastoma 1
Msmb	Microseminoprotein, beta	Fsh	Follicle-stimulating hormone
Pdlim3	PDZ and LIM domain 3	Notch1	NOTCH1
Klf4	Kruppel-like factor 4	H2afx	H2A histone family, member X
Sirt1	Sirtuin 1	Runx2	Runt-related transcription factor 2
Atf3	Activating transcription factor 3	Psen1	Presenilin 1
Bcl2	B-cell CLL/lymphoma 2	Dhfr	Dihydrofolate reductase
Mcm3	Minichromosome maintenance complex component 3	Tgfbr2	Transforming growth factor, beta receptor II
Met	Mesenchymal-epithelial transition factor	Ptpre	Protein tyrosine phosphatase, receptor type, E
Ppara	Peroxisome proliferator-activated receptor alpha	Creb	cAMP response element-binding protein
Tsc1	Tuberous sclerosis 1	Ikbip	IKBKB interacting protein
Birc5	Baculoviral IAP repeat containing 5	Chuk	Conserved helix-loop-helix ubiquitous kinase
Il-8	Interleukin 8	Ctgf	Connective tissue growth factor
Bdnf	Brain-derived neurotrophic factor	Cfl1	Cofilin 1
Acta2	F-actin	Cg	Cancer-germline
Mapre2	Microtubule-associated protein, RP/EB family, member 2		

Table A3 Differentially regulated genes of PDGFRα-positive cells at different time points. Length (bp) of transcript isoforms, log2 (fold change) of FPKM and q-values are shown. Colors indicate the log2 (FC) of transcript expression level up-regulated at P5, W6 or in NC; blue = up-regulated at W6, orange = up-regulated at P5, green = up-regulated in NC. 999.0 = infinity (transcript not expressed in one condition) bp = base pair, FC = fold change, NC = negative control, W6 = week 6, P5 = postnatal day 5. "-" = not differentially regulated.

Gene	Length (bp)	log2(FC) W6 vs. P5 (q value)	log2(FC) W6 vs. NC (q value)	log2(FC) P5 vs. NC (q value)
Pdgfrα	6553	2.7 (0.002)	2.2 (0.001)	-
Pdgfrl	1548	3.8 (0.008)	-	-2.0 (0.008)
Acta2	1781	3.39 (0.002)	2.07 (0.001)	-1.36 (0.008)
Eln	3950	3.15 (0.002)	-	-
Myh11	6593	3.89 (0.002)	2.06 (0.007)	-1.87 (0.02)
Myl6	665	-	1.54 (0.04)	2.24 (0.01)
Myl7	590	1.70 (0.02)	-	-2.94 (0.02)
Myl9	1128	3.32 (0.002)	2.79 (0.001)	-
Cotl1	1520	-1.50 (0.002)	-3.46 (0.001)	-1.99 (0.002)
Il17ra	3970	-2.02 (0.002)	-3.46 (0.001)	-1.47 (0.006)
Pparg	1857	-2.24 (0.03)	-3.02 (0.01)	-
Rxra	4905	-1.46 (0.01)	-	-
Lpl	1900	-2.70 (0.002)	-3.05 (0.001)	-
Thy1	1735	-3.17 (0.002)	-4.18 (0.001)	-
Slc27a6	2485	4.09 (0.002)	-	-2.93 (0.002)
Spry1	2451	2.65 (0.002)	2.39 (0.003)	-
Spry4	4587	2.34 (0.002)	2.90 (0.001)	-
Col15a1	5172	1.61 (0.02)	-	-
Col1a2	5346	2.57 (0.002)	2.30 (0.02)	-
I Col1a2	5318	-	-	-999.00 (0.03)
Col12a1	11511	2.74 (0.002)	1.96 (0.001)	-
Col3a1	5564	2.17 (0.02)	2.89 (0.03)	-
Col5a2	6625	2.63 (0.002)	2.85 (0.001)	-
Col4a4	10018	1.94 (0.003)	-	-
Col1a1	5933	2.76 (0.01)	-	-
Col6a2	3949	2.36 (0.002)	2.25 (0.001)	-

Col4a5	6536	-	2.47 (0.01)	-
Col6a1	4100	1.78 (0.002)	2.26 (0.001)	-
Col23a1	5658	2.08 (0.002)	2.12 (0.04)	-
Ctgf	2398	1.80 (0.002)	1.92 (0.001)	-
Smo	3779	2.72 (0.002)	2.86 (0.001)	-
Tgfbi	2678	1.18 (0.02)	-4.03 (0.001)	-5.24 (0.002)
Tgfb2	4725	4.15 (0.02)	-	-
Tgfb3	3384	3.30 (0.01)	2.94 (0.05)	-
Tgfbr1	5756	-1.91 (0.002)	-1.56 (0.01)	-
Tgfbr2	8013	-1.79 (0.04)	-	-
Tgfbr3	6087	1.55 (0.002)	1.67 (0.001)	-
Tgfa	3762	-	1.94 (0.01)	1.81 (0.05)
Ltbp2	6769	4.28 (0.002)	2.29 (0.001)	-2.03 (0.01)
Ltbp3	5190	2.75 (0.002)	1.77 (0.01)	-
Ltbp4	5118	1.54 (0.002)	2.81 (0.01)	-
Mir-24-2	9585	-	-1.47 (0.02)	-1.31 (0.03)
Mir-487b	452	-	-	-999.00 (0.01)
Mir-382	477	-	-	-999.00 (0.02)
Mir-29a	10026	-3.50 (0.002)	-3.79 (0.001)	-
Mir-30a	10175	1.58 (0.05)	-	-

Table A4 Genes up-regulated in PDGFRα-positive cells at P5 vs. NC and W6 vs. NC, and those of them differentially regulated between P5 vs. W6. The length (bp) of transcript isoforms, log2 (fold change) of FPKM and q-values are shown. Positive numbers indicate the log2 (FC) of transcript expression level up-regulated at P5 and negative numbers at W6; bp = base pair, FC = fold change, W6 = week 6, P5 = postnatal day 5, NC = negative control.

Gene	Length (bp)	log2(FC) W6 vs. P5	q value
Stfa1	439	4.1	0.02
H19	2286	1.99	0.000
Kntc1	6961	1.32	0.000
Per2	5835	1.19	0.04
Tgfbi	2678	1.18	0.02
Hist1h2bj	463	1.02	0.02
Golm1	3557	-1.04	0.01

Lcn2	915	-1.07	0.01
Parp8	3088	-1.08	0.01
Lrp12	4095	-1.17	0.000
Mfsd7c	3435	-1.18	0.000
Tnfrsf21	3626	-1.24	0.000
Ramp1	2549	-1.32	0.000
Gm12867	12926	-1.34	0.01
Cx3cr1	3729	-1.4	0.000
Cotl1	1520	-1.5	0.000
Itga4	9767	-1.71	0.000
Pygl	2825	-1.77	0.000
Il1a	1974	-1.92	0.000
Itgal	5194	-1.96	0.000
Sirpa	3637	-1.98	0.000
Fpr1	1324	-1.98	0.000
Il17ra	3970	-2.02	0.000
Csf2rb	4764	-2.03	0.000
1600029D21Rik	1846	-2.04	0.000
Fcer1g	674	-2.06	0.000
Plcg2	4296	-2.15	0.000
Syk	5122	-2.21	0.01
Lsp1	1395	-2.22	0.000
Ccl6	1440	-2.31	0.000
Tlr2	3013	-2.31	0.000
Ccdc21	3128	-2.34	0.000
Mrc1	5352	-2.34	0.000
Pld4	1978	-2.34	0.000
Marco	1985	-2.35	0.000
Cyfip2	6298	-2.35	0.000
Stk10	5026	-2.36	0.000
Nckap1l	4717	-2.36	0.000
Rac2	3036	-2.36	0.000
Ctss	1355	-2.37	0.000
Tyrobp	564	-2.4	0.000
Mpeg1	4335	-2.41	0.000
Cybb	4752	-2.46	0.000

Ccdc88b	4959	-2.5	0.000
1810033B17Rik	1132	-2.61	0.000
Ptafr	3637	-2.62	0.000
Cxcl2	1080	-2.65	0.000
Ear1	713	-2.65	0.000
Atp6v0d2	2539	-2.69	0.000
Itgb2	2970	-2.71	0.000
Ccl6	776	-2.72	0.000
Cd53	2802	-2.75	0.000
Clec7a	2205	-2.79	0.000
Niacr1	1938	-2.82	0.000
AI504432	8892	-2.82	0.000
Card11	4112	-2.83	0.000
Dock2	6409	-2.92	0.000
Il1b	1356	-2.96	0.000
S100a8	392	-2.98	0.000
Ptprc	4609	-3.01	0.000
Cxcl2	1830	-3.06	0.000
Cd14	1572	-3.11	0.000
Gp49a	5396	-3.3	0.000
Trem1	3024	-3.45	0.000

Table A5 Gene sets enrichment for canonical pathways analysis for W6 time point. Normalized enrichment score (NES) and family-wise error rate (FWER) values are shown. GSEA was used, references: [85].

Canonical Pathway	NES	FWER value
REACTOME: Chemokine receptors bind chemokines	2.59	0.000
REACTOME: Cytokine signaling in immune system	2.57	0.000
KEGG: Hematopoietic cell lineage	2.56	0.000
REACTOME: Costimulation by the CD28 family	2.55	0.000
REACTOME: Signaling by ILs	2.53	0.000
REACTOME: Influenza viral RNA transcription and replication	2.52	0.000
REACTOME: Translation	2.52	0.000
REACTOME: Peptide chain elongation	2.48	0.000
REACTOME: SRP dependent cotranslational protein targeting to	2.48	0.000

membrane		
REACTOME: Influenza life cycle	2.46	0.001
REACTOME: 3' UTR mediated translational regulation	2.46	0.001
PID: IL12 STAT4 pathway	2.46	0.001
KEGG: Lysosome	2.45	0.001
REACTOME: Nonsense mediated decay enhanced by the exon junction complex	2.44	0.001
PID: BCR 5 pathway	2.44	0.001
KEGG: Intestinal immune network for IGA production	2.44	0.001
KEGG: Ribosome	2.43	0.001
PID: NFAT TF pathway	2.41	0.001
REACTOME: Toll receptor cascades	2.39	0.001
REACTOME: TCR signaling	2.35	0.002
REACTOME: Interferon alpha beta signaling	2.34	0.002
KEGG: NOD-like receptor signaling pathway	2.34	0.002
BIOCARTA: NO2 IL12 pathway	2.33	0.005
KEGG: Apoptosis	2.32	0.005
KEGG: Leishmania infection	2.32	0.005
KEGG: JAK STAT signaling pathway	2.32	0.005
REACTOME: Metabolism of mRNA	2.31	0.005
REACTOME: Generation of second messenger molecules	2.31	0.005
BIOCARTA: CTLA4 pathway	2.31	0.005
BIOCARTA: T helper pathway	2.30	0.005
PID: PI3KCI pathway	2.29	0.005
PID: CD8 TCR downstream pathway	2.28	0.006
REACTOME: Nucleotide binding domain leucine rich repeat containing receptor NLR signaling pathway	2.26	0.006
REACTOME: Signaling by the B cell receptor BCR	2.25	0.007
REACTOME: Metabolism of RNA	2.25	0.008
BIOCARTA: T cytotoxic pathway	2.24	0.008
KEGG: Antigen processing and presentation	2.23	0.009
PID: HIV NEF pathway	2.23	0.009
REACTOME: Interferon signaling	2.22	0.009
ST: T cell signal transduction	2.21	0.011
KEGG: FC epsilon ri signaling pathway	2.19	0.018
REACTOME: Activated TLR4 signaling	2.19	0.018

BIOCARTA: Stem pathway	2.19	0.019
KEGG: Graft versus host disease	2.18	0.021
REACTOME: Adaptive immune system	2.18	0.021
KEGG: RIG I-like receptor signaling pathway	2.16	0.027
REACTOME: PD1 signaling	2.15	0.029
REACTOME: Interferon gamma signaling	2.15	0.029
SIG: BCR signaling pathway	2.15	0.029
REACTOME: Activation of the mRNA upon binding of the Cap binding complex and EIFS and subsequent binding to 43S	2.15	0.031
BIOCARTA: NKT pathway	2.13	0.042
REACTOME: GPVI-mediated activation cascade	2.13	0.042
BIOCARTA: DC pathway	2.13	0.043
PID: REG GR pathway	2.13	0.044
PID: IL4 2 pathway	2.12	0.046
REACTOME: MYD88 MAL cascade initiated on plasma membrane	2.12	0.047
REACTOME: IL- 3, -5 and GM-CSF signaling	2.12	0.049

LIST OF PUBLICATIONS

1. **K. Ahlbrecht, F. Klein, D. Dontireddy, S. Becker, S. Bellusci, W.D. Richardson, M. Szibor, T. Braun, R.E. Morty, W. Seeger, and R. Voswinckel.** *Pulmonary PDGFRα positive cells derive multiple mesenchymal cell lineages in different tissue compartments of the lung.* Submitted.
2. **D. Dontireddy, K. Ahlbrecht, S. Günther, M. Looso, C. Künne, , W. Bertrams, B. Schmeck, T. Braun, R.E. Morty, W. Seeger, R. Voswinckel,** *Identification of pulmonary PDGFRα-positive fibroblast miRNome and transcriptome during postnatal lung development.* In preparation.

BIBLIOGRAPHY

1. Ambros, V., *microRNAs: Tiny regulators with great potential.* Cell, 2001. **107**(7): p. 823-826.
2. Artemenko, Y., et al., *Anti-adipogenic effect of PDGF is reversed by PKC inhibition.* Journal of Cellular Physiology, 2005. **204**(2): p. 646-653.
3. Awonusonu, F., et al., *Developmental shift in the relative percentages of lung fibroblast subsets: role of apoptosis postseptation.* American Journal of Physiology-Lung Cellular and Molecular Physiology, 1999. **277**(4): p. L848-L859.
4. Balamugesh, T. *Pulmonary Interstitium*. Pulmonary & Critical Care Bulletin 2001. **VII**.
5. Barkauskas, C.E., et al., *Type 2 alveolar cells are stem cells in adult lung.* Journal of Clinical Investigation, 2013. **123**(7): p. 3025-3036.
6. Bartel, D.P., *MicroRNAs: Genomics, biogenesis, mechanism, and function.* Cell, 2004. **116**(2): p. 281-297.
7. Bartel, D.P., *MicroRNAs: Target Recognition and Regulatory Functions.* Cell, 2009. **136**(2): p. 215-233.
8. Basciani, S., et al., *Expression of platelet-derived growth factor (PDGF) in the epididymis and analysis of the epididymal development in PDGF-A, PDGF-B, and PDGF receptor beta deficient mice.* Biology of Reproduction, 2004. **70**(1): p. 168-177.
9. Betsholtz, C., *Biology of platelet-derived growth factors in development.* Birth Defects Research Part C: Embryo Today: Reviews, 2003. **69**(4): p. 272-285.
10. Betsholtz, C., L. Karlsson, and P. Lindahl, *Developmental roles of platelet-derived growth factors.* Bioessays, 2001. **23**(6): p. 494-507.
11. Bhaskaran, M., et al., *MicroRNA-127 modulates fetal lung development.* Physiological Genomics, 2009. **37**(3): p. 268-278.
12. Blervaque, R. *Technology Overview sequencing Ion Torrent PGM*. 2013; Available from: http://www.biorigami.com/?tag=report-ion-torrent.
13. Bostrom, H., A. Gritli-Linde, and C. Betsholtz, *PDGF-A/PDGF alpha-receptor signaling is required for lung growth and the formation of alveoli but not for early lung branching morphogenesis.* Developmental dynamics : an official publication of the American Association of Anatomists, 2002. **223**(1): p. 155-62.
14. Bostrom, H., et al., *PDGF-A signaling is a critical event in lung alveolar myofibroblast development and alveogenesis.* Cell, 1996. **85**(6): p. 863-873.
15. Boucherat, O., et al., *Gene expression profiling in lung fibroblasts reveals new players in alveolarization.* Physiol Genomics, 2007. **32**(1): p. 128-41.
16. Boucherat, O., et al., *Gene expression profiling in lung fibroblasts reveals new players in alveolarization.* Physiological Genomics, 2007. **32**(1): p. 128-141.
17. Bruce, M.C., C.E. Honaker, and R.J. Cross, *Lung fibroblasts undergo apoptosis following alveolarization.* American Journal of Respiratory Cell and Molecular Biology, 1999. **20**(2): p. 228-236.
18. Burgess, H.A., et al., *PPAR gamma agonists inhibit TGF-beta induced pulmonary myofibroblast differentiation and collagen production: implications for therapy of lung fibrosis.* American Journal of Physiology-Lung Cellular and Molecular Physiology, 2005. **288**(6): p. L1146-L1153.
19. Burri, P.H., *Structural aspects of postnatal lung development - alveolar formation and growth.* Biol Neonate, 2006. **89**(4): p. 313-22.
20. Casaletto, J.B. and A.I. McClatchey, *Spatial regulation of receptor tyrosine kinases in development and cancer.* Nature Reviews Cancer, 2012. **12**(6): p. 386-399.
21. Cheloufi, S., et al., *A Dicer-independent miRNA biogenesis pathway that requires Ago catalysis.* Nature, 2010. **465**(7298): p. 584-U76.

22. Chen, L.L., et al., *Dynamic Regulation of Platelet-Derived Growth Factor Receptor alpha Expression in Alveolar Fibroblasts during Realveolarization.* American Journal of Respiratory Cell and Molecular Biology, 2012. **47**(4): p. 517-527.
23. Chu, Y.J. and D.R. Corey, *RNA Sequencing: Platform Selection, Experimental Design, and Data Interpretation.* Nucleic Acid Therapeutics, 2012. **22**(4): p. 271-274.
24. Cock, P.J., et al., *The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants.* Nucleic Acids Res, 2010. **38**(6): p. 1767-71.
25. Corporation, L.T., *Ion Proton™ System for Next-Generation Sequencing*. 2013.
26. Croce, C.M. and G.A. Calin, *miRNAs, cancer, and stem cell division.* Cell, 2005. **122**(1): p. 6-7.
27. Czech, B. and G.J. Hannon, *Small RNA sorting: matchmaking for Argonautes.* Nature Reviews Genetics, 2011. **12**(1): p. 19-31.
28. Del Fabbro, C., et al., *An Extensive Evaluation of Read Trimming Effects on Illumina NGS Data Analysis.* PLoS One, 2013. **8**(12): p. e85024.
29. Deutsch GH, P.H., *Prenatal Lung Development*, in *Chronic Obstructive Lung Disease*, V.N.a.M. W, Editor. 2002, BC Decker Inc.: London.
30. Diederichs, S. and D.A. Haber, *Dual role for argonautes in MicroRNA processing and Posttranscriptional regulation of MicroRNA expression.* Cell, 2007. **131**(6): p. 1097-1108.
31. Dong, J., et al., *MicroRNA Networks in Mouse Lung Organogenesis.* PLoS One, 2010. **5**(5).
32. Doucet, J., et al., *Molecular cloning and functional characterization of mouse coactosin-like protein.* Biochemical and Biophysical Research Communications, 2002. **290**(2): p. 783-789.
33. Du, B., et al., *High glucose down-regulates miR-29a to increase collagen IV production in HK-2 cells.* Febs Letters, 2010. **584**(4): p. 811-816.
34. Dueck, A. and G. Meister, *MicroRNA processing without Dicer*. Genome Biology, 2010. **11**(6).
35. Ebert, M.S. and P.A. Sharp, *Roles for microRNAs in conferring robustness to biological processes.* Cell, 2012. **149**(3): p. 515-24.
36. Eden, E., et al., *GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists.* BMC Bioinformatics, 2009. **10**.
37. Exiqon. *What are microRNAs?* ; Available from: http://www.exiqon.com/what-are-microRNAs.
38. Ezzie, M.E., et al., *Gene expression networks in COPD: microRNA and mRNA regulation.* Thorax, 2012. **67**(2): p. 122-131.
39. Gan, Q., et al., *Smooth muscle cells and myofibroblasts use distinct transcriptional mechanisms for smooth muscle alpha-actin expression.* Circulation Research, 2007. **101**(9): p. 883-892.
40. Guo, L. and Z.H. Lu, *Global expression analysis of miRNA gene cluster and family based on isomiRs from deep sequencing data*. Computational Biology and Chemistry, 2010. **34**(3): p. 165-171.
41. Hall, N., *Advanced sequencing technologies and their wider impact in microbiology.* Journal of Experimental Biology, 2007. **210**(9): p. 1518-1525.
42. Hamilton, T.G., et al., *Evolutionary divergence of platelet-derived growth factor alpha receptor signaling mechanisms.* Mol Cell Biol, 2003. **23**(11): p. 4013-25.
43. Hashimoto, S., et al., *Expression of Spred and Sprouty in developing rat lung.* Gene Expr Patterns, 2002. **2**(3-4): p. 347-53.
44. Hawkins, S.M., et al., *Functional MicroRNA Involved in Endometriosis.* Molecular Endocrinology, 2011. **25**(5): p. 821-832.
45. He, J., et al., *miRNA-Mediated Functional Changes through Co-Regulating Function Related Genes.* PLoS One, 2010. **5**(10).

46. Heid, H.W., et al., *Adipophilin is a specific marker of lipid accumulation in diverse cell types and diseases.* Cell and Tissue Research, 1998. **294**(2): p. 309-321.
47. Hinz, B. and G. Gabbiani, *Fibrosis: recent advances in myofibroblast biology and new therapeutic perspectives.* F1000 biology reports, 2010. **2**: p. 78.
48. Hinz, B., et al., *The myofibroblast - One function, multiple origins.* American Journal of Pathology, 2007. **170**(6): p. 1807-1816.
49. Hoch, R.V. and P. Soriano, *Roles of PDGF in animal development.* Development, 2003. **130**(20): p. 4769-4784.
50. Hsu, S.D., et al., *miRTarBase: a database curates experimentally validated microRNA-target interactions.* Nucleic Acids Research, 2011. **39**(Database issue): p. D163-9.
51. Iwasaki, S., et al., *Hsc70/Hsp90 Chaperone Machinery Mediates ATP-Dependent RISC Loading of Small RNA Duplexes.* Molecular Cell, 2010. **39**(2): p. 292-299.
52. J.M., P. *SANGER WHO? Sequencing the next generation*. AAAS/SCIENCE, 2009. DOI: 10.1126/science.opms.p0900033.
53. Kaplan, N.B., M.M. Grant, and J.S. Brody, *The Lipid Interstitial Cell of the Pulmonary Alveolus - Age and Species-Differences.* American Review of Respiratory Disease, 1985. **132**(6): p. 1307-1312.
54. Kimani, P.W., et al., *PDGF-R alpha gene expression predicts proliferation, but PDGF-A suppresses transdifferentiation of neonatal mouse lung myofibroblasts.* Respiratory Research, 2009. **10**.
55. Kimani, P.W., et al., *PDGF-Ralpha gene expression predicts proliferation, but PDGF-A suppresses transdifferentiation of neonatal mouse lung myofibroblasts.* Respir Res, 2009. **10**: p. 119.
56. Kimura, J. and G.H. Deutsch, *Key mechanisms of early lung development.* Pediatr Dev Pathol, 2007. **10**(5): p. 335-47.
57. Koumas, L., et al., *Thy-1 expression in human fibroblast subsets defines myofibroblastic or lipofibroblastic phenotypes.* American Journal of Pathology, 2003. **163**(4): p. 1291-1300.
58. Koumas, L., T.J. Smith, and R.P. Phipps, *Fibroblast subsets in the human orbit: Thy-1(+) and Thy-1(-) subpopulations exhibit distinct phenotypes.* European Journal of Immunology, 2002. **32**(2): p. 477-485.
59. Kozomara, A. and S. Griffiths-Jones, *miRBase: integrating microRNA annotation and deep-sequencing data.* Nucleic Acids Research, 2011. **39**: p. D152-D157.
60. Kriegel, A.J., et al., *MiR-382 targeting of kallikrein 5 contributes to renal inner medullary interstitial fibrosis.* Physiol Genomics, 2012. **44**(4): p. 259-67.
61. Krol, J., I. Loedige, and W. Filipowicz, *The widespread regulation of microRNA biogenesis, function and decay.* Nature Reviews Genetics, 2010. **11**(9): p. 597-610.
62. Kulkarni, A.A., et al., *PPAR-gamma Ligands Repress TGF beta-Induced Myofibroblast Differentiation by Targeting the PI3K/Akt Pathway: Implications for Therapy of Fibrosis.* PLoS One, 2011. **6**(1).
63. Laczny, C., et al., *miRTrail - a comprehensive webserver for analyzing gene and miRNA patterns to enhance the understanding of regulatory mechanisms in diseases.* BMC Bioinformatics, 2012. **13**.
64. Lakatos, H.F., et al., *The Role of PPARs in Lung Fibrosis.* Ppar Research, 2007.
65. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biology, 2009. **10**(3).
66. Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C-Elegans Heterochronic Gene Lin-4 Encodes Small Rnas with Antisense Complementarity to Lin-14.* Cell, 1993. **75**(5): p. 843-854.
67. Leslie, K.O., J. Mitchell, and R. Low, *Lung Myofibroblasts.* Cell Motility and the Cytoskeleton, 1992. **22**(2): p. 92-98.

68. Li, J. and G.W. Hoyle, *Overexpression of PDGF-A in the lung epithelium of transgenic mice produces a lethal phenotype associated with hyperplasia of mesenchymal cells.* Developmental Biology, 2001. **239**(2): p. 338-349.
69. Li, Z.Y., et al., *Biological Functions of miR-29b Contribute to Positive Regulation of Osteoblast Differentiation.* Journal of Biological Chemistry, 2009. **284**(23): p. 15676-15684.
70. Lin, J.G. and A.P. Chen, *Activation of peroxisome proliferator-activated receptor-gamma by curcumin blocks the signaling pathways for PDGF and EGF in hepatic stellate cells.* Laboratory Investigation, 2008. **88**(5): p. 529-540.
71. Lindahl, P., et al., *Alveogenesis failure in PDGF-A-deficient mice is coupled to lack of distal spreading of alveolar smooth muscle cell progenitors during lung development.* Development, 1997. **124**(20): p. 3943-3953.
72. Liu, B., J. Li, and M.J. Cairns, *Identifying miRNAs, targets and functions.* Briefings in Bioinformatics, 2012.
73. Lo, S.S., et al., *Overexpression of miR-370 and downregulation of its novel target TGF beta-RII contribute to the progression of gastric carcinoma.* Oncogene, 2012. **31**(2): p. 226-237.
74. Londos, C., et al., *Role of PAT proteins in lipid metabolism.* Biochimie, 2005. **87**(1): p. 45-49.
75. Luna, C., et al., *Cross-talk between miR-29 and Transforming Growth Factor-Betas in Trabecular Meshwork Cells.* Investigative Ophthalmology & Visual Science, 2011. **52**(6): p. 3567-3572.
76. Mardis, E.R., *The impact of next-generation sequencing technology on genetics.* Trends in Genetics, 2008. **24**(3): p. 133-141.
77. Massaro, G.D., D. Massaro, and P. Chambon, *Retinoic acid receptor-alpha regulates pulmonary alveolus formation in mice after, but not during, perinatal period.* American Journal of Physiology-Lung Cellular and Molecular Physiology, 2003. **284**(2): p. L431-L433.
78. McGowan, S., et al., *Mice bearing deletions of retinoic acid receptors demonstrate reduced lung elastin and alveolar numbers.* American Journal of Respiratory Cell and Molecular Biology, 2000. **23**(2): p. 162-167.
79. McGowan, S.E., et al., *Platelet-Derived Growth Factor Receptor-Alpha-Expressing Cells Localize to the Alveolar Entry Ring and Have Characteristics of Myofibroblasts During Pulmonary Alveolar Septal Formation.* Anatomical Record-Advances in Integrative Anatomy and Evolutionary Biology, 2008. **291**(12): p. 1649-1661.
80. McGowan, S.E. and D.M. McCoy, *Fibroblasts Expressing PDGF-Receptor-Alpha Diminish During Alveolar Septal Thinning in Mice.* Pediatric Research, 2011. **70**(1): p. 44-49.
81. McGowan, S.E. and J.S. Torday, *The pulmonary lipofibroblast (lipid interstitial cell) and its contributions to alveolar development.* Annual Review of Physiology, 1997. **59**: p. 43-62.
82. Medicine, T.J.o.G., *Gene Therapy Clinical Trials Database*. J. Gene Med.
83. Milosevic, J., et al., *Profibrotic Role of miR-154 in Pulmonary Fibrosis.* American Journal of Respiratory Cell and Molecular Biology, 2012. **47**(6): p. 879-887.
84. Mizuno, S., et al., *MicroRNA-199a-5p Is Associated With Hypoxia-Inducible Factor-1 alpha Expression in Lungs From Patients With COPD.* Chest, 2012. **142**(3): p. 663-672.
85. Mootha, V.K., et al., *PGC-1 alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.* Nature Genetics, 2003. **34**(3): p. 267-273.
86. Muzumdar, M.D., et al., *A global double-fluorescent cre reporter mouse.* Genesis, 2007. **45**(9): p. 593-605.

87. NCBI. *Entrez Gene: TGFBI transforming growth factor, beta-induced, 68kDa*. Available from: http://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=ShowDetailView&TermToSearch=7045.
88. Ohare, K.H. and M.N. Sheridan, *Electron Microscopic Observations on Morphogenesis of Albino Rat Lung, with Special Reference to Pulmonary Epithelial Cells.* American Journal of Anatomy, 1970. **127**(2): p. 181-&.
89. Ott, C.E., et al., *MicroRNAs Differentially Expressed in Postnatal Aortic Development Downregulate Elastin via 3 ' UTR and Coding-Sequence Binding Sites.* PLoS One, 2011. **6**(1).
90. Papadopoulos, G.L., et al., *DIANA-mirPath: Integrating human and mouse microRNAs in pathways.* Bioinformatics, 2009. **25**(15): p. 1991-1993.
91. Pasquinelli, A.E., et al., *Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA.* Nature, 2000. **408**(6808): p. 86-89.
92. Post, M. and I. Copland, *Overview of lung development.* Acta Pharmacologica Sinica, 2002. **23**: p. 4-7.
93. Provost, P., et al., *Coactosin-like protein, a human F-actin-binding protein: critical role lysine-75.* Biochemical Journal, 2001. **359**: p. 255-263.
94. Rehan, V.K., et al., *Evidence for the presence of lipofibroblasts in human lung.* Experimental Lung Research, 2006. **32**(8): p. 379-393.
95. Robenek, H., et al., *Spatial integration of TIP47 and adipophilin in macrophage lipid bodies.* Journal of Biological Chemistry, 2005. **280**(7): p. 5789-5794.
96. Rothberg, J.M., et al., *An integrated semiconductor device enabling non-optical genome sequencing.* Nature, 2011. **475**(7356): p. 348-352.
97. Rusk, N., *Torrents of sequence.* Nature Methods, 2011. **8**(1): p. 44-44.
98. Sakurai, R., et al., *1alpha,25(OH)2D3 and its 3-epimer promote rat lung alveolar epithelial-mesenchymal interactions and inhibit lipofibroblast apoptosis.* Am J Physiol Lung Cell Mol Physiol, 2009. **297**(3): p. L496-505.
99. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA Sequencing with Chain-Terminating Inhibitors.* Proceedings of the National Academy of Sciences of the United States of America, 1977. **74**(12): p. 5463-5467.
100. Sasson, A. and T.P. Michael, *Filtering error from SOLiD Output.* Bioinformatics, 2010. **26**(6): p. 849-50.
101. Sayed, D. and M. Abdellatif, *MicroRNAs in development and disease.* Physiol Rev, 2011. **91**(3): p. 827-87.
102. Schittny, J.C., et al., *Programmed cell death contributes to postnatal lung development.* American Journal of Respiratory Cell and Molecular Biology, 1998. **18**(6): p. 786-793.
103. Schittny, J.C., S.I. Mund, and M. Stampanoni, *Evidence and structural mechanism for late lung alveolarization.* American Journal of Physiology-Lung Cellular and Molecular Physiology, 2008. **294**(2): p. L246-L254.
104. Schultz, C.J., et al., *Role of adipocyte differentiation-related protein in surfactant phospholipid synthesis by type II cells.* American Journal of Physiology-Lung Cellular and Molecular Physiology, 2002. **283**(2): p. L288-L296.
105. Sessa, R., Hata, A., *Role of microRNAs in lung development and pulmonary diseases.* Pulm Circ 2013(3): p. 315-28.
106. Shi, W., S. Bellusci, and D. Warburton, *Lung development and adult lung diseases.* Chest, 2007. **132**(2): p. 651-6.
107. Shirdel, E.A., et al., *NAViGaTing the Micronome - Using Multiple MicroRNA Prediction Databases to Identify Signalling Pathway-Associated MicroRNAs.* PLoS One, 2011. **6**(2).
108. Simon, D.M. and T.J. Mariani, *Role of PPARs and Retinoid X Receptors in the Regulation of Lung Maturation and Development.* Ppar Research, 2007.

109. Smith, L.M., et al., *Fluorescence detection in automated DNA sequence analysis.* Nature, 1986. **321**(6071): p. 674-9.
110. Sun, T., et al., *A human YAC transgene rescues craniofacial and neural tube development in PDGFR alpha knockout mice and uncovers a role for PDGFR alpha in prenatal lung growth.* Development, 2000. **127**(21): p. 4519-4529.
111. Tennis, M.A., et al., *Sprouty-4 Inhibits Transformed Cell Growth, Migration and Invasion, and Epithelial-Mesenchymal Transition, and Is Regulated by Wnt7A through PPAR(Y) in Non-Small Cell Lung Cancer.* Molecular Cancer Research, 2010. **8**(6): p. 833-843.
112. Tomankova, T., M. Petrek, and E. Kriegova, *Involvement of microRNAs in physiological and pathological processes in the lung.* Respiratory Research, 2010. **11**.
113. Torday, J.S. and V.K. Rehan, *The evolutionary continuum from lung development to homeostasis and repair.* American Journal of Physiology-Lung Cellular and Molecular Physiology, 2007. **292**(3): p. L608-L611.
114. Torday, J.S., E. Torres, and V.K. Rehan, *The role of fibroblast transdifferentiation in lung epithelial cell proliferation, differentiation, and repair in vitro.* Pediatr Pathol Mol Med, 2003. **22**(3): p. 189-207.
115. Trapnell, C., L. Pachter, and S.L. Salzberg, *TopHat: discovering splice junctions with RNA-Seq.* Bioinformatics, 2009. **25**(9): p. 1105-1111.
116. Trapnell, C., et al., *Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.* Nature Biotechnology, 2010. **28**(5): p. 511-U174.
117. Van Pottelberge, G.R., et al., *MicroRNA Expression in Induced Sputum of Smokers and Patients with Chronic Obstructive Pulmonary Disease.* American Journal of Respiratory and Critical Care Medicine, 2011. **183**(7): p. 898-906.
118. van Rooij, E., et al., *Dysregulation of microRNAs after myocardial infarction reveals a role of miR-29 in cardiac fibrosis.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(35): p. 13027-13032.
119. Varisco, B.M., et al., *Thy-1 Signals through PPAR gamma to Promote Lipofibroblast Differentiation in the Developing Lung.* American Journal of Respiratory Cell and Molecular Biology, 2012. **46**(6): p. 765-772.
120. Wang, F.E., et al., *MicroRNA-204/211 alters epithelial physiology.* Faseb Journal, 2010. **24**(5): p. 1552-1571.
121. Watts, J.K. and D.R. Corey, *Silencing disease genes in the laboratory and the clinic.* J Pathol, 2012. **226**(2): p. 365-79.
122. Wienholds, E. and R.H.A. Plasterk, *MicroRNA function in animal development.* Febs Letters, 2005. **579**(26): p. 5911-5922.
123. Williams, A.E., et al., *MicroRNA Expression Profiling in Mild Asthmatic Human Airways and Effect of Corticosteroid Therapy.* PLoS One, 2009. **4**(6).
124. Wirth, A., et al., *G12-G13-LARG-mediated signaling in vascular smooth muscle is required for salt-induced hypertension.* Nat Med, 2008. **14**(1): p. 64-8.
125. Xi, S., et al., *Cigarette smoke mediates epigenetic repression of miR-487b during pulmonary carcinogenesis.* J Clin Invest, 2013. **123**(3): p. 1241-61.
126. Xin, H.B., et al., *Smooth muscle expression of Cre recombinase and eGFP in transgenic mice.* Physiological Genomics, 2002. **10**(3): p. 211-215.
127. Yamada, M., et al., *Temporal expression of alpha-smooth muscle actin and drebrin in septal interstitial cells during alveolar maturation.* Journal of Histochemistry & Cytochemistry, 2005. **53**(6): p. 735-744.
128. Yang, J.S., et al., *Widespread regulatory activity of vertebrate microRNA* species.* Rna-a Publication of the Rna Society, 2011. **17**(2): p. 312-326.
129. Yang, S.Z., et al., *miR-145 regulates myofibroblast differentiation and lung fibrosis.* Faseb Journal, 2013. **27**(6): p. 2382-2391.

130. Ying, S.Y., *MicroRNA Protocols*. 2006: Humana Press.

DECLARATION

„Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne unzulässige Hilfe oder Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nichtveröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten sowie ethische, datenschutzrechtliche und tierschutzrechtliche Grundsätze befolgt. Ich versichere, dass Dritte von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen, oder habe diese nachstehend spezifiziert. Die vorgelegte Arbeit wurde weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zweck einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt. Alles aus anderen Quellen und von anderen Personen übernommene Material, das in der Arbeit verwendet wurde oder auf das direkt Bezug genommen wird, wurde als solches kenntlich gemacht. Insbesondere wurden alle Personen genannt, die direkt und indirekt an der Entstehung der vorliegenden Arbeit beteiligt waren. Mit der Überprüfung meiner Arbeit durch eine Plagiatserkennungssoftware bzw. ein internetbasiertes Softwareprogramm erkläre ich mich einverstanden.“

Gießen , 2014 Daria Dontireddy

Ort, Datum Unterschrift

ACKNOWLEDGMENT

My Journey of Doctoral study and subsequent dissertation has taken more than three and half years, which I can recall as hard but enjoyable. Much has happened and changed in meantime. This time was a tangle of many queries and simultaneously strengthening self-confidence. It would be difficult to go through it without the support of many wonderful people, especially my mother, friends and dearest husband. I would like to take this opportunity to acknowledge their advice, support and inspiration throughout this doctoral time otherwise most important time of my life.

I'd like to give a special thanks to Prof. Dr. med. Werner Seeger for offering me the opportunity to work on this project and attend to the international graduate program Molecular Biology and Medicine of the Lung (MBML)

I'd also like to give a special thanks to P.D. Dr. med. Robert Voswinckel for his constructive ideas and discussions. His constant support and for given opportunities to collaborate with other scientists without whom this work would not be created.

Thirdly, my heartfelt thanks go to Dr. med. Katrin Ahlbrecht for her personal cheering and relaxed demeanor which helped to build good working relationship.

My special thanks go to Dr. Rory Morty whose knowledge and experience made the MBML program so useful and effective. My gratitude is also extended to Prof. Dr. med. Bernd Schmeck and his group members, especially Dr. Wilhelm Bertrams whose in depth technical knowledge helped me to understand complex phenomenons.

My special thanks go to Dr. Stefan Günther, Dr. Mario Looso and Dr. Carsten Künne for their professional collaboration and helpful suggestions. I'd like to thank my all lab colleagues and especially friends: Célimène, Anita, Rebecca and Minmin for all those enjoyable moments not only in the lab but also outside the work. Thanks girls!

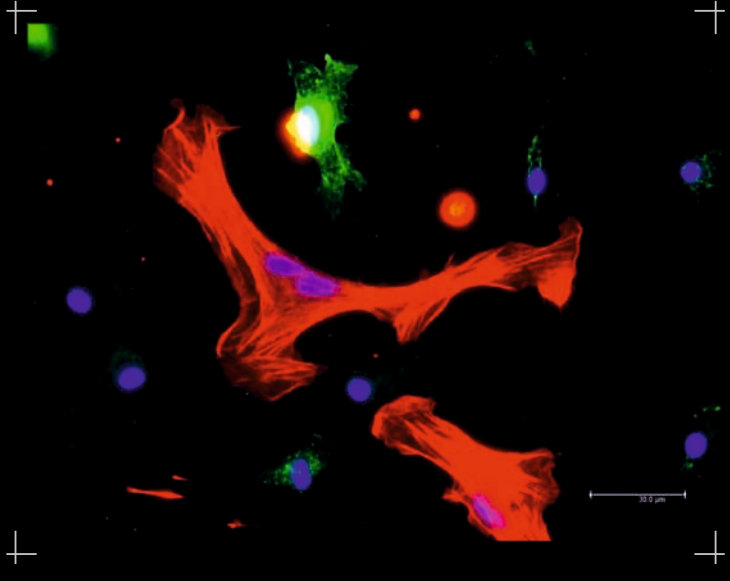
Of course no acknowledgments would be complete without giving thanks to my beloved mother. She taught me about hard work and self-respect, about persistence and about how to be independent.

Last, but not least, thanks go to my beloved husband Rakesh, who supported me during this journey and withstood over three and half years of separation. Thank you for giving me your love, and your heart.

This work has been supported by the Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz (LOEWE).

Der Lebenslauf wurde aus der elektronischen Version der Arbeit entfernt.

The curriculum vitae was removed from the electronic version of the paper.



édition scientifique
VVB LAUFERSWEILER VERLAG

VVB LAUFERSWEILER VERLAG
STAUFENBERGRING 15
D-35396 GIESSEN

Tel: 0641-5599888 Fax: -5599890
redaktion@doktorverlag.de
www.doktorverlag.de

ISBN: 978-3-8359-6306-1