NTNU
Norwegian University of
Science and Technology

MicroRNA, environmental carcinogens and risk of lung cancer

Audun Trygge Haugen Bersaas

Biotechnology
Submission date: May 2011
Supervisor: Aage Haugen, IBT
Co-supervisor: Steen Mollerup, STAMI

Norwegian University of Science and Technology
Department of Biotechnology

Acknowledgements

The work for this thesis was carried out at the National Institute of Occupational Health, Oslo, as part of the Master of Science in Biotechnology at the Norwegian University of Science and Technology. I would like to thank my supervisors, Professor Aage Haugen, and PhD Steen Mollerup. Especially, I want to thank Steen Mollerup for exceedingly good guidance through the work for this thesis, and for his commitment and enthusiasm for the project.

I wish to thank Nur Duale at the Norwegian Institute of Public Health for all his help with data analysis. A special thank to Rita Bæra and Tove Andreassen for invaluable tutoring, and patience, in the laboratory. I would also like to thank everyone at the group of toxicology.

Thanks to all my fellow students at the top floor, Torunn Rønningen, Lone Nilsen, Aqsa Mahmood, Mari Sjøberg, and Kjærsti Johnsen for all good times at the office.

Last, I want to thank my family for always supporting me.

Oslo, 16.05.2011

Audun Trygge Haugen Bersaas

Abstract

Lung cancer is the leading cause of cancer related deaths world wide. The major cause of lung cancer is tobacco smoking. Tobacco smoke contains many different carcinogens including polycyclic aromatic hydrocarbons (PAH), which are known to cause DNA damage and initiate tumourigenesis.

The aryl hydrocarbon receptor (AHR) is a basic helix-loop-helix protein of the PER-ARNT-SIM (PAS) family of transcription factors, which is involved in many biological processes including development, cell cycle regulation, cell differentiation, cell-cell signalling, cellular growth, and circadian rhythms. Most of the known function of AHR, however, is regarding its role in transcriptional induction of xenobiotic metabolising enzymes (XMEs). In the lung, PAHs bind and activate the AHR, leading to transcriptional induction of members of the cytochrome P450 (*CYP*) enzyme family that are involved in activation of pro-carcinogenic PAHs to carcinogenic metabolites.

MicroRNAs (miRNAs), a group of short non-coding RNA molecules, mediate sequence specific silencing of messenger RNAs (mRNAs) and are important in regulation of gene expression. The deregulation of certain miRNAs has been observed in several types of cancer, including lung cancer. Like mRNAs, miRNAs belong to the class II type of genes, and transcription of these types of RNA can therefore be regulated by the same transcription factors.

Little is known regarding the role of miRNAs in environmental toxicology. The aim of this thesis was to identify miRNAs regulated by the AHR, and consequently may be important in mediating toxicity of tobacco smoke carcinogens.

Expression levels of 750 miRNAs in mouse lung of *Ahr* wild-type and knockout animals were profiled by RT-qPCR arrays. Fifteen miRNAs were found to be differentially expressed between the groups, of which only one was up-regulated. The biological pathways in which these miRNAs may be involved were studied by an bioinformatic approach. Four miRNAs from the mouse experiment, whose expression were most significantly divergent, together with three miRNAs previously reported to be regulated by the *Ahr* in murine cell lines, were studied in an RNA interference (RNAi) experiment in immortalised human lung epithelial cell lines. Two

of these miRNAs, both selected from the literature, were found to have altered expression in AHR knockdown cells compared to control cells. Discrepancies between AHR regulated miRNAs identified in the *in vivo* and *in vitro* experiments were observed. Taken together this suggest a difference between effect of the AHR in vivo and *in vitro*, rather than differences between species.

In conclusion, it was found that expression of several miRNAs possibly is regulated in an AHR-dependent manner in lung. Discrepancies between *in vivo* and *in vitro* experiments suggest that care must be taken when extrapolating miRNA expression data from cell culture studies to whole organs or organisms.

Abbreviations

3-MC	3-methylcolantrene
3‘UTR	3’ untranslated region
ADC	Adenocarcinoma
AGO	Argonaute
AHR	Aryl hydrocarbon receptor
AHRE	AHR response element
AIP	AHR interacting protein
ANOVA	Analysis of variance
ARNT	AHR nuclear transferase
B[*a*]P	Benzo[*a*]pyrene
bHLH	Basic helix-loop-helix
bp	Base pair
BPDE	(±)-anti-7β-8α-dihydroxy-9α,10α-epoxy-7,8,9,10-tetrahydro-benzo[a]p
cAMP	Cyclic adenosine monophosphate
CCND	Cyclin D1
CDK4	Cyclin dependent kinase 4
cDNA	Complementary DNA
C_q	Quantification cycle
CSC	Cigarette smoke condensate
CYP	Cytochrome P450
DEP	Diesel exhaust particles
DGCR8	DiGeorge syndrome critical region gene 8
DLC	Dioxin-like compounds
DNMT	DNA methyltransferase
dsRNA	Double stranded RNA
ECM	Extra cellular matrix
ECS	Environmental cigarette smoke
EGFR	Epidermal growth factor receptor
ESR1	Oestrogen receptor 1
GOI	Gene of interest
GPCR	G-protein coupled receptor
GRP	Gastrin releasing peptide
HBEC3-KT	Human bronchial epithelial cells
hsa	*Homo sapiens*
HSP90	Heat shock protein 90 kDa
LCC	Large cell carcinoma

LDA	Low-density array
miRISC	miRNA containing RISC
miRNA	microRNA
mmu	*Mus musculus*
mRNA	Messenger RNA
NF	Normalisation factor
NNK	4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone
NNN	N'-nitrosonornicotine
NSCLC	Non-small cell lung cancer
nt	Nucleotides
NTC	Non-template control
PAH	Polycyclic aromatic hydrocarbon
PAS	PER-ARNT-SIM
PCA	Principal component analysis
PM	Particulate matter
pre-miRNA	Precursor-miRNA
pri-miRNA	Primary pre-miRNA
Rb	Retinoblastoma
RISC	RNA induced silencing complex
RNAi	RNA interference
RT-qPCR	Reverse transcriptase quantitative polymerase chain reaction
SCC	Squamous cell carcinoma
SCLC	Small cell lung cancer
siRISC	siRNA containing RISC
siRNA	Short interfering RNA
SNP	Single nucleotide polymorphism
TCDD	Tetrachlorodibenzo-p-dioxin
TERT	Telomerase reverse transcriptase
TP-23	HSP90 co-chaperon
TRBP	TAR RNA binding protein
TSNA	Tobacco specific nitrosamines
TSS	Transcription start site
XME	Xenobiotic metabolising enzymes
XPO5	Exportin-5
XRE	Xenobiotic response element (same as AHRE)

•

Content

1. Introduction

1.1. Lung cancer

Cancer is the leading cause of death in the economically developed world, and the second leading cause of death in the economically developing world. In 2008, approximately 12.7 million new cases of cancer were diagnosed, and 7.6 million cancer related deaths were recorded world wide. The rate of cancer is increasing in economically developing countries as a result of population growth and ageing, as well as an increasing adaption of cancer associated lifestyle choices. This includes smoking, physical inactivity, and "westernised" diets[1,2].

Globally lung cancer is the second most frequently diagnosed cancer and the leading cause of cancer death for females in both economically developed and developing countries, only bypassed by breast cancer. For males, lung cancer is the most frequently diagnosed cancer in developing countries, and the second most diagnosed

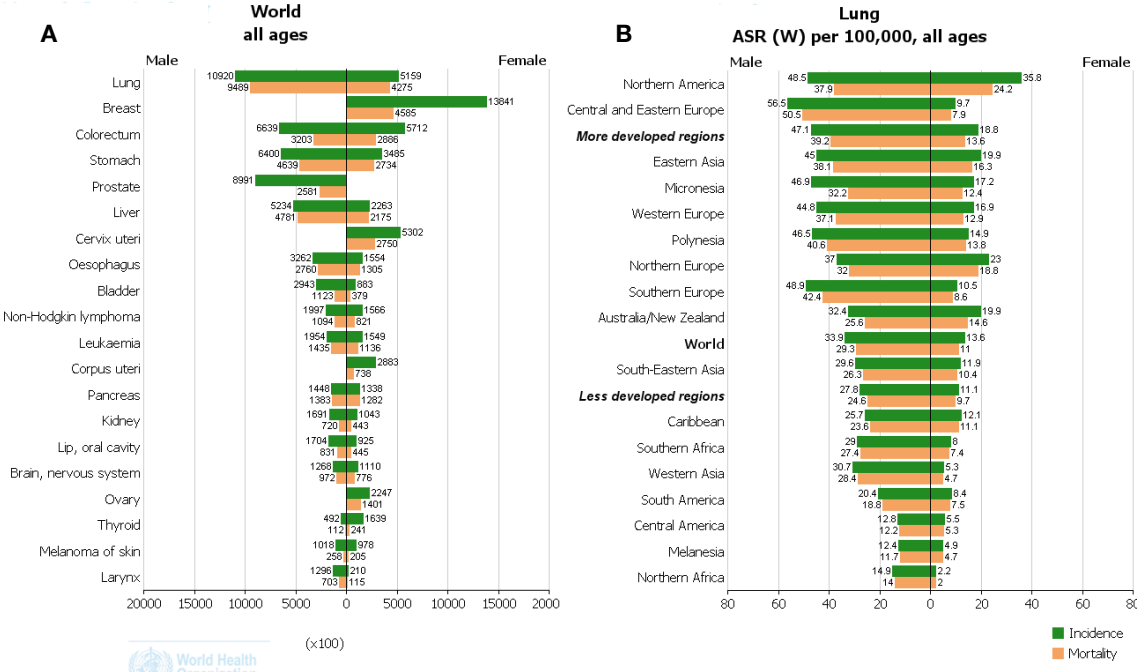


Figure 1.1: A) Incidence and mortality rates the 20 most common types of cancer in male and female. **B)** Incidence and mortality rates of lung cancer in males and females in different regions of the world[2].

cancer in developed countries (Figure 1.1).

Due to its strong association with tobacco smoking, lung cancer death rates are decreasing in most western countries, where tobacco epidemic peaked by the middle of the last century. In contrast, lung cancer rates are increasing in economically developing countries, for example in China and several other countries in Asia, and in several African countries. This is probably linked to increasing economical status enabling citizens to buy tobacco more recently in these countries[1].

Lung cancer can be divided into two main types: Small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). SCLCs are the most aggressive forms of the disease and have a greater potential to metastasise than other type of lung cancer. NSCLCs are a group of different types of cancers, including squamous cell carcinoma (SCC), adenocarcinoma (ADC), and large-cell carcinoma (LCC) along with other minor subtypes[3]. Cigarette smoking is significantly associated with all histological types of lung cancer[4], including adenocarcinoma, which also is the most common cancer type in never smokers[3].

1.1.1. Risk factors

Tobacco smoking is strongly associated with lung cancer, and may also play a significant role in cancers of other sites like larynx, nasal cavity, oral cavity, oesophagus, liver, pancreas, cervix, bladder, and leukaemia[5]. Tobacco smoking accounts for 80-90% of the worldwide lung cancer burden in males, and for at least 50% in females[1].

The strong association between lung cancer and cigarette smoking suggests that environmental carcinogens cause most cases of lung cancer. A carcinogen is any agent, chemical, physical, or viral, that causes cancer or increases the incidence of cancer. Cigarette smoke contains many different carcinogens, the most important, based on carcinogenic potency and levels in smoke are considered polycyclic aromatic hydrocarbons (PAH), N-nitrosamines, aromatic amines, aldehydes, ethylene epoxides, 1,3-butadiene, and benzene (Table 1.1), of which the two last mentioned compounds are the most prevalent strong carcinogens[5]. N-nitrosamines are a large class of carcinogens that affect different tissues depending on their structure. The two most studied nitrosamines are 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) and N'-nitrosonornicotine (NNN), both only found in tobacco, and therefore termed tobacco specific nitrosamines (TSNA)[5].

Risk of lung cancer is strongly associated with duration of smoking, and numbers of cigarettes smoked

also influence the relative risk. Smoking cessation is associated with reduction of relative risk of lung cancer among all smokers. This reduction in risk is observed within 1-4 years of cessation in both men and women, and the magnitude of the reduction in relative risk increases with increased time since cessation[4].

Other risk factors for lung cancer include cooking fumes, radon, asbestos, arsenic, and PAHs from environmental pollution[1]. An increased risk of lung cancer has also been associated with certain occupations, such as transport workers, and in food processing and food service occupations[6]. Exposure to indoor radon is estimated to be responsible for 9% of all lung cancer deaths in Europe[3].

Urban air particulate matter (PM) has been associated with increased risk of lung cancer[7,8]. PM is a complex mixture of airborne solid and liquid particles including soot, organic material, sulphates, nitrates, other salts, metals, and biological material. Carcinogenic compounds, such as PAH, are able to adsorb to the particles and may be deposited in the respiratory tract[8]. The combustion of fossil fuels is the source of most of the organic and inorganic compounds, oxidants, and acids, and contributes heavily to particulate air pollution in most urban settings[9].

Table 1.1: Some of the major groups of tobacco carcinogens and the most importent within each group. Modified from IARC Monograph 83 (2004)[4]

Agent	Notable carcinogens	Amount in mainstream cigarette smoke
Polycyclic aromatic hydrocarbons	Benzo[*a*]pyrene	8.5-11.6 ng
N-Nitrosamines	NNK	110-133 ng
	NNN	154-196 ng
Aromatic amines	4-Aminobiphenyl	2-5 ng
	2-Naphthylamine	1-22 ng
Aldehydes	Formaldehyde	10.3-25 µg
	Acetaldehyde	770-864 µg
Volatile hydrocarbons	1,3-butadiene	20-40 µg
	Benzene	12-50 µg
Miscellaneous organic compounds	Ethylene oxide	7 µg
	Vinyl chloride	11-15 µg

1.1.2. Genetic susceptibility to Lung Cancer

Even though there is a strong relationship between cigarette smoking and lung cancer, only 10-15% of all smokers develop lung cancer[10]. This, besides the cases of some non-smokers developing lung cancer, suggests that there might be differences in individual predisposition to lung cancer. There is also evidence for a 2-fold in-

creased risk of lung cancer for individuals with a family history of lung cancer strongly indicating a heritable influence on susceptibility of lung cancer[11].

Predispositions in "hallmark of cancer" genes

Tumourigenesis is a multistep process characterised by accumulation of mutations resulting in defects in regulatory circuits that govern normal cell proliferation and homeostasis, and, ultimately, leads to invasive malignancy[12,13]. In their transformation from normal to malignant state, cancer cells need to gain at least eight essential changes known as the “hallmarks of cancer”, first defined by Hanahan and Weinberg in 2000[13]. The six essential alterations that cancer cells must gain are: Sustaining proliferative signals, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating tissue invasion and metastasis. Recently, the same authors added two new emerging hallmarks; deregulation of cellular energetics and avoiding immune destruction. In addition, two enabling characteristics that drive tumour progression were included: genome instability and mutation, and tumour-promoting inflammation (Figure 1.2)[14].

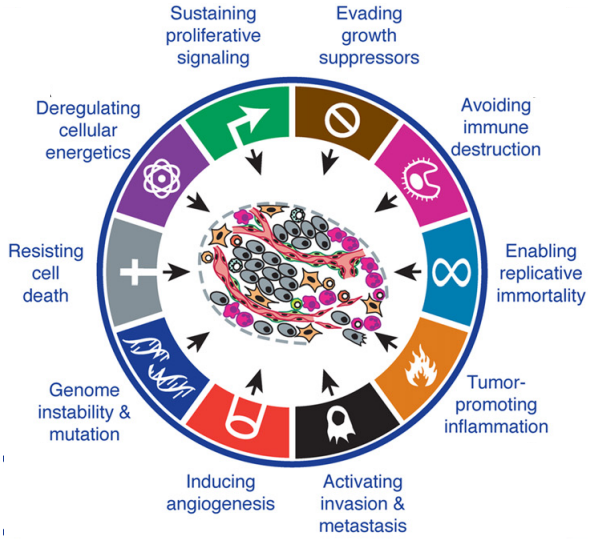


Figure 1.2: Hallmarks of cancer.
Eight characteristics that cells need to attain to become mutagenic, and two characteristics that enable tumour cells to grow. From Hanahan & Weinberg (2011)[14]

The eight hallmarks of cancer each describe anticancer defence mechanisms that function to avoid and eliminate abnormalities, and cancer cells needs to prevail these safety mechanism to gain malignancy[13]. Inter-individual differences in susceptibility to cancer can be due to single nucleotide polymorphisms (SNPs) in genes involved in any of these defence mechanisms, including DNA repair genes, telomerase genes, cell cycle regulating genes, or genes involved in inflammation[15-17].

For example, a mutation in epidermal growth factor receptor (*EGFR*) has been associated with NSCLC, and mutations in the tumour suppressor gene *TP53* in families with

Li-Fraumeni syndrome are associated with a slightly higher frequency of lung cancer and a lower age of occurrence[15]. The latter gene is activated by DNA damage, hypoxia, or aberrant oncogene expression and can subsequently promote DNA repair and apoptosis, and hence important in the defence mechanism against cancer[18]. Polymorphism in inflammation associated genes may also play a role in cancer predisposition. For instance, a polymorphism in the pro-inflammatory genes interleukin-1 beta (*IL1B*) and *COX2* has been associated with increased risk of non-small cell lung cancer[19,20].

Predisposition in xenobiotic metabolising and DNA repair genes

In relation to predisposition to lung cancer, SNPs in genes involved in metabolism of exogenous carcinogens are of importance. Metabolism of xenobiotics is divided into three phases, each containing different types of xenobiotic metabolising enzymes (XMEs) that metabolise, detoxifies, and excrete exogenous compounds. The first phase involves functionalisation, executed by several different classes of enzymes, including mixed function oxidase (cytochrome P450 or CYP), flavin monooxygenase, alcohol dehydrogenase, and epoxide hydrolase. Phase II XMEs are responsible for conjugation, either of compounds directly from the environment, or from metabolites of phase I enzymes. Important phase II metabolising enzymes are UDP-glycosynyl transferase (UGT), sulphotransferase, and glutathione S-transferase (GST). Phase III enzymes are responsible for further metabolism of phase II conjugates, such as metabolism of glutathione conjugates by gamma-glutanyl-transpeptidase. Phase III enzymes are also responsible for the transport of the conjugates out of the cells and into bile or urine[21].

Polymorphisms resulting in altered activity of any of the enzymes involved in xenobiotic metabolism can potentially lead to an increased risk of cancer, as a result of increased metabolic activation, decreased conjugation, or decreased excretion[22]. For instance, an association between early onset (age < 51) lung cancer in female and specific polymorphisms in cytochrome P4501B1 (*CYP1B1*) and *CYP2A13* has been shown[23]

Polymorphisms in DNA repair genes may cause interidividual differences in DNA repair capacity. Metabolites of pro-carcinogens can bind DNA and form adducts, and if left unrepaired, these adducts can cause mutations that eventually may initiate tumour formation. An SNP in the DNA repair gene Xerodema Pigmentosum (*XP*) C, has been associated with a longer time to progression of NSCLC[24]

There is evidence of differences in lung cancer susceptibility between males and females, both in epidemiological studies and in laboratory studies[25]. Higher level of DNA-adducts have been found in females compared to that of males, both *in vivo* and *in vitro*[26,27]. The higher DNA-adduct level in females may result from significantly higher levels of CYP1A1 and thus increased bioactivation of certain tobacco carcinogens. One possible explanation for this sex difference could be related to the AHR, but the authors did not find any differences in expression levels of AHR mRNA[26]. The difference in susceptability to lung cancer in male and female suggest that hormonal factors may be involved in lung tumourigenesis[25,28].

1.1.3. Gene environment interactions

Only few cancers are caused by either environmental exposure or the individuals genetic make-up. Instead, the majority of cancers, including lung cancer, are caused by an interplay between environmental and genetic factors, so-called gene-environment interactions.

Gene-environment interactions can be divided into two categories: 1) diseases resulting from interactions between the individual genetic make-up and environmental factors, 2) diseases resulting from direct effects of environmental exposure on the genome, leading to for example DNA damage and/or mutations[29].

In any type of cancer, there are often significant differences in age of onset, rapidity of tumour growth, presence of metastases, pathological appearance, gene expression patterns, somatic genetic changes, response to therapy, and familial risk[30]. Identification of gene-environment interactions in tobacco-related cancers have proven to be complex due to several reasons, including the large number of genes involved in tobacco metabolism, the numerous substrates metabolised by relevant genes and the interaction of smoking with other metabolic pathways[22].

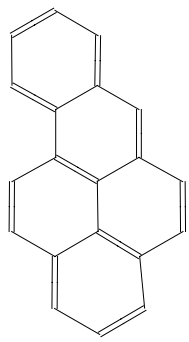
1.2.The aryl hydrocarbon receptor

1.2.1.AHR basics

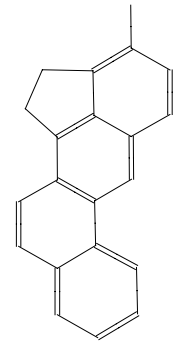
The AHR was first discovered in 1976 by Poland et al. who discovered that tetrachlorodibenzo-p-dioxin (TCDD or simply dioxin) binds specifically to a receptor that induces expression of Cytochrome P450 1A1 (*CYP1A1*) and UDP-glucosynyl transferase in mouse liver[31]. The AHR can bind planar polycyclic ligands, for instance PAH and dioxins and dioxin-like compounds (DLC). The two most studied AHR ligands are the PAH benzo[*a*]pyrene

(B[*a*]P), and TCDD[32]. Also some dietary chemicals (for example flavonoids, carotenoids, and phenolics) can function as AHR ligands and activate transcription of AHR targeted genes. However, these compounds seem to be relatively weak AHR ligands. Conversion of dietary indoles, including tryptophan, in the mammalian digestive tract to significantly more potent AHR ligands has been demonstrated. There has also been an ongoing search for one or more endogenous AHR ligands, but no conclusions have been made. UV photoproducts of tryptophan and histidine can directly bind to and activate AHR[32,33].

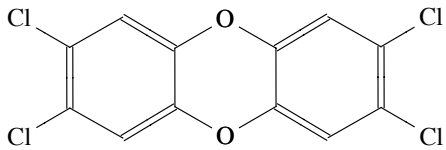
The AHR is a member of the basic helix-loop-helix/PER ARNT Sim (bHLH/PAS) family of transcriptional regulators, which have important roles in adaptive responses to generalised and cellular stress, besides playing key developmental roles. The bHLH motif is common to a wide range of transcription factors all of which bind to DNA as dimers. bHLH/PAS proteins are separated into two classes based on their



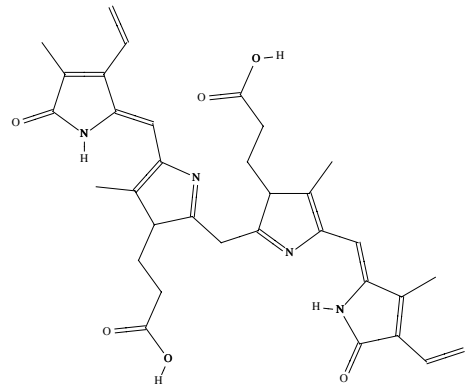
Benzo[*a*]pyrene (B[*a*]P)



3-Methylcolantrene (3-MC)



TCDD (Dioxin)



Bilirubin

Figure 1.3: Chemical structure of some known ligands of the aryl hydrocarbon receptor (AHR). Benzo[*a*]pyrene and Dioxin is the two most extensively studied.

dimerisation potential. Class I factors are only capable of hetero-dimerisation with class II factors which, in contrast, are capable of both hetero- and homo-dimerisation. Class I factors are located in the cytoplasm and functions as a sensory unit, that upon stimulation are translocated to the nucleus where they dimerise with class II factors, which resides in the nucleus. Upon dimerisation, the bHLH/PAS complex is able to bind DNA and activate transcription of target genes [34]. The AHR, a bHLH/PAS class I protein, is the only vertebrate member of the bHLH/PAS family known to bind and be activated by small chemical ligands[34].

1.2.2. Activation of AHR

In the unliganded, inactive state, the AHR is mainly located in the cytoplasm bound to a homodimer of the 90 kDa heat shock protein (HSP90), one molecule of AHR interacting protein (AIP), and one HSP90 co-chaperon TP-23 (Figure 1.4). When a ligand associates with the AHR, a conformational change reveals a nuclear localisation signal in the AHR, which is recognised by importin-β. The entire AHR/HSP90/AIP/TP-23 complex is translocated to the nucleus, where the chaperon protein complex is exchanged by the AHR nuclear translocator (ARNT) protein[35]. After dimerisation, the AHR/ARNT complex recognises and binds DNA at specific sequences in the promoter region or in the enhancer of its target genes. The AHR/ARNT complex can recognise and bind to two different response elements, termed AH response element I (AHRE-I, also called xenobiotic response element [XRE], or dioxin responsive element [DRE])[36] and AHRE-II[37].

The activity of the exogenous ligand activated AHR can be inhibited by G-protein signalling pathways. Oesch-Bartlomowicz *et al.* (2005) found that when Hepa-1 cells were stimulated with cyclic-AMP (cAMP), AHR was activated and translocated to the nucleus. However, the cAMP-activated AHR were not able to induce transcription of CYP1A1. Moreover, cAMP reduced the effect of TCDD-activated AHR[38]. Nakata *et al.* (2009) found that G-proteins negatively regulates AHR activity by competitively sequestering AIP from the AHR complex. This leads to ubiquitination of AHR by an unknown E3 ubiquitin ligase, and ultimately results in translocation of AHR to the nucleus. As for the cAMP-activated AHR, transcription of AHR/ARNT target genes after exposure to 3-MC was inhibited by G-protein activated AHR[39].

1.2.3. AHR in cellular processes

Originally AHR was studied as a pure xenobiotic metabolism activator. It has become evident, however, that the

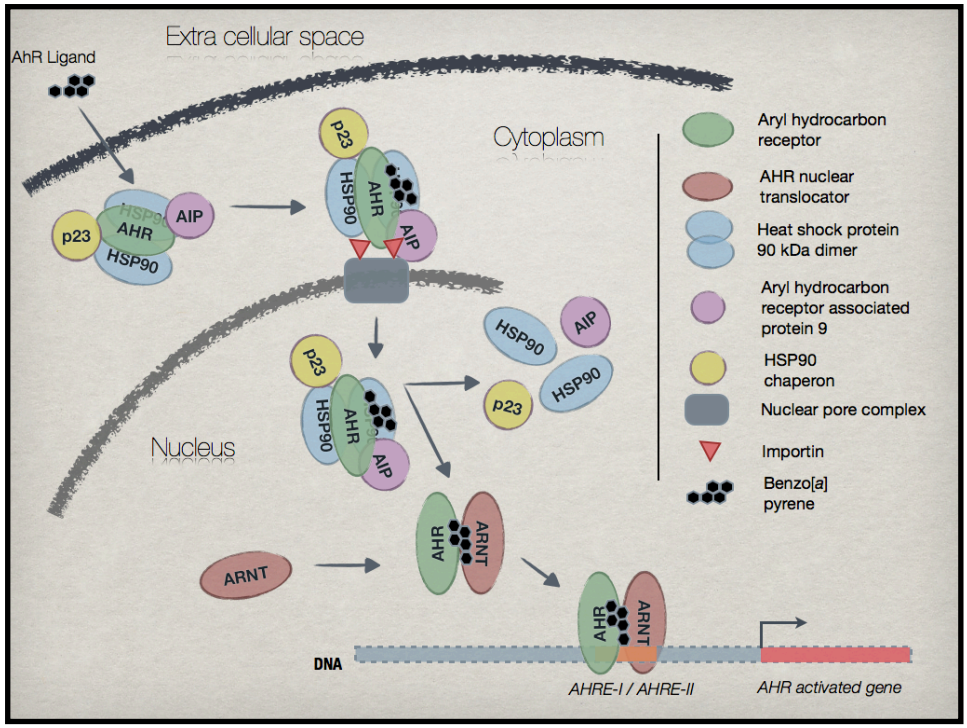


Figure 1.4: Unliganded AHR is located in the cytoplasm complexed with a HSP90 dimer, AIP and HSP90 co-chaperon p23. Upon ligand binding the complex translocates to the nucleus, where the chaperon proteins are exchanged for ARNT binding to AHR. The AHR/ARNT heterodimer then binds response elements (AHRE-I or AHRE-II) in promoters or enhancers of target genes and induces transcription. Modified from Furness *et al.* (2007)[32].

AHR is involved in a vast range of biological processes, including development, cell cycle regulation, cell signalling, cytoskeleton organisation, extra cellular matrix (ECM) modification, adhesion, immune responses, and circadian rhythm[34,40-43].

Although most of its known actions are dependent on ligand binding, the AHR seems to be able to activate some biological processes independent of an exogenous ligand. However, it is not known if endogenous ligands are involved in these processes since no such ligand has been identified[44]. Interestingly, AHR in a transcriptionally active form has been found to be localised in the nucleus of cells in low confluence culture, and as the confluence increases, AHR is translocated to the cytoplasm, a process associated with Ca2+ concentration[45]. This suggests a connection between cell-cell signalling mechanisms and AHR functions.

Barhoover *et al*. (2010) showed that the AHR forms a complex with CDK4 and CCND1 in the absence of an exogenous ligand. This complex can phosphorylate RB1, leading to G1 to S phase cell cycle transition. Expo-

sure to an exogenous ligand will, however, result in dissociation of AHR from CDK4 and CCND1, resulting in decreased RB1 phosphorylation and G1 cell cycle arrest[46]. Thus it seems that AHR may function as an environmental sensor that allows certain biological processes, such as cell cycle progression, only to occur in the absence of toxic compounds[47].

In contrast to PAHs, TCDD is metabolically inert and hence does not lead to genotoxic metabolites. However, TCDD is known to be a potent tumour promoter, suggesting that tumours are formed in another way than through DNA-adduct formation[41]. Tijet *et al*. (2006) showed that expression of 456 genes was altered by TCDD in an AHR-dependent manner, while only expression of 32 genes were altered in an AHR-independent manner[48]. In addition they found that 392 genes were significantly differentially expressed between *Ahr* knockout ($Ahr^{-/-}$) and wildtype ($Ahr^{+/+}$) mice wihtout TCDD exposure. This demonstrates that most, if not all, of the tumourigenic effects of TCDD are mediated through AHR, and that carcinogenicity of AHR ligands is not solely mediated by direct DNA damage.

1.2.4. Metabolism and carcinogenesis of PAH

One of the most extensively studied groups of carcinogens in relation to lung cancer are the PAHs. These compounds are products of incomplete combustion and occur as mixtures in cigarette smoke, tars, soots, broiled foods, adsorbed to particulate matter, automobile engine exhaust (particularly diesel), and other materials generated by incomplete combustion[5]. PAHs are procarcinogens that are activated to carcinogens by XMEs in the body. In phase I, PAHs are metabolised primarily by CYP1A1 or CYP1B1, but in human lung CYP1A1 is the most important enzyme in the bioactivation of B[*a*]P[49]. These activated, carcinogenic metabolites can form adducts with cellular macromolecules and disrupt normal cell function if left unrepaired[50].

The most widely studied PAH is benzo[*a*]pyrene (B[*a*]P). This PAH is not carcinogenic in itself, but in the body it is activated in a three step process to the carcinogenic metabolite (±)-anti-7β-8α-dihydroxy-9α,10α-epoxy-7,8,9,10-tetrahydro-benzo[*a*]pyrene (BPDE), which can form DNA-adducts at the N2 position of deoxyguanosine (BPDE-N2-dG). The first step in B[*a*]P metabolism is catalysed by CYP enzymes, forming (7R,8S)-epoxy-7,8-dihydrobenzo[a]pyrene (B[*a*]P-oxide). This is converted to (7R,8R)-dihydroxy-7,8-dihydrobenzo[a]pyrene (B[*a*]P-diol), catalysed by epoxide hydrolase. The last step is catalysed by CYP and other

enzymes, producing mainly BPDE (three other 7,8-diol-9,10-epoxide isomers of B[*a*]P are also formed, but to a lesser extent than BPDE)[51].

Unrepaired DNA-adducts in tumour suppressor genes or proto-oncogenes and consequently mutations in these genes may prompt tumorigenesis. Adducts from B[*a*]P metabolites mainly result in G→T or G→A mutations, depending on the DNA sequence[50]. BPDE DNA-adducts have been observed in the *TP53* tumour suppressor gene, which is mutated 60% of all lung cancers[52].The oncogenic *KRAS* gene is mutated in approximately 30%-40% of adenocarcinoma[53]. In Chinese women, there has been found an association between the use of smokey cole for heating and cooking, and risk of lung cancer. The emissions from the coal consist of 43% PAH. In lung tumours studied by DeMarini *et al*., the *KRAS* gene was found to be mutated in 29% of the cases, with mainly G→T transversions[54].

1.3.Epigenetics in Cancer

Epigenetics is defined as somatically heritable changes in gene expression that are not a consequence of changes in the DNA sequence. Epigenetics comprise a number of mechanisms, including DNA methylation, histone modification, and microRNAs (miRNAs). Animal studies indicates that prenatal and early postnatal environmental factors, including nutritional supplements, xenobiotic chemicals, behavioural cues, reproductive factors, and low-dose radiation, can result in altered epigenetic programming and subsequent changes in the risk of developing disease[55]. Epigenetic changes can be inherited mitotically in somatic cells, which indicates that environmental effects on the epigenome can result in long-term effects on gene expression[55].

To better understand gene-environment interactions, epigenetics should be added to classic genetic analysis. Changes in DNA methylation pattern, chromatin structure, or miRNA expression. Polymorphism in miRNAs or in miRNA-targeted sequences in 3'UTR of mRNAs, may also result in phenotypic changes. For example, polymorphisms in the 3'UTR of ATP binding cassette (*ABC*) *B1* and *ABCC1* which may alter miRNA binding, have been associated with an increased risk of lung cance[56].

As mentioned previously, gene-environment interactions can be divided into two groups (see section 1.2). In the same way, epigenetic effect can be divided into two similar groups: 1) diseases resulting from epigenetic differences making individuals less or more vulnerable to environmental in-

sults, 2) diseases resulting from alteration of epigenetic states caused by environmental exposure, for example changes in histone modification, DNA methylation, or miRNA expression[29]. Epigenetic alterations might be inherited transgenerationally, thereby potentially affecting the health of future generations. However, although epigenetic states of some genomic regions have proven to be inherited across generations in animal models, no such inheritance has so far been proven in humans[29,55].

DNA methylation

DNA methylation is a covalent modification, heritable by somatic cells after cell division, and is important in regulating many cellular processes, including chromatin structure, X-chromosome inactivation, genomic imprinting, chromosome stability, and gene transcription[57]. DNA can be methylated at the 5-position of a cytosine ring in a CpG dinucleotide. CpG sites are asymmetrically distributed into CpG poor regions and CpG dense regions called CpG islands. These CpG islands are often located in the promoter region of protein coding genes, and their methylation is associated with gene silencing. Aberrant DNA methylation of tumor suppressor genes occurs in many types of cancer.

DNA methylation is carried out by DNA methylatransferases (DNMTs). There are three major DNMTs with different specific tasks; DNMT3A and 3B are responsible for de novo methylation, while DNMT1 is responsible for maintaining already existing methylation patterns[58]. In cancer cells, CpG islands that are normally unmethylated may become methylated, or conversely, some genes that are normally methylated may become unmethylated, resulting in activation of normally silenced genes, such as oncogenes or retrotransposons[59].

Histone modification

DNA is packed into chromatin, a dynamic structure consisting of 146 base pairs (bp) of DNA wrapped around nucleosome protein complexes. The nucleosome is composed of an octamer of four core histones, an H3/H4 tetramer and two H2A/H2B dimers[59]. Histones can be modified by addition of chemical groups to their tails. Types of modification include acetylation, methylation, phosphorylation, sumoylation, and ADP ribosylation, resulting in opening or condensing of the chromatine structure.

The most common modifications are acetylation and methylation of lysine residues in the amino terminal of histone 3 (H3) and histone 4 (H4). Increased acetylation induces transcription activation, whereas decreased acetylation usually induces transcription repression. Methylation of histones

is associated with repression or activation of transcription, depending on the lysine residue position[29]. Besides modification of histones, nucleosomes can be modified by incorporation of different histone variants. For instance, the histone variant macroH2A have been shown to be involved in regulation of *HOXA* gene cluster in zebrafish development, and in X-chromosome inactivation in females[60].

RNA interference

RNA interference (RNAi) conducted by double-stranded (ds)RNA molecules was first described in 1998 by Fire and co-workers[61], who discovered that injection of short dsRNA (20-30 nucleotide [nt]) into *Caenorhabtidis elegans* resulted in phenotypes similar to that expected in knock out animals. They found that these phenotypes resulted from specific knock down of mRNAs containing sequences complimentary to the injected siRNA. For their discovery, Andrew Fire and Craig Mello were awarded the Nobel Prize in Physiology or Medicine in 2006, less than ten years after their original discovery.

The effects of small RNAs on gene expression are generally inhibitory, and the corresponding regulatory mechanisms are therefore collectively termed RNA silencing. The central theme is that the small RNAs serve as specificity factors that direct associated effector proteins to target nucleic acid molecules via base-pairing interactions. Invariably, the core component of the effector machinery is a member of the Argonaute protein superfamily. Small RNAs are grouped into three main categories; short interfering RNAs (siRNA), microRNAs (miRNA), and piwi-interacting RNAs (piRNA), based on their origins, structures, associated effector proteins, and biological roles[62].

1.4. miRNA

The first discovery of small RNA molecules involved in the cells innate mechanism for regulation of gene expression came in 1993 when Ambros, Feinbaum, and Lee found an RNA-molecule coded by the *lin-4* gene to be important in regulation of the *lin-14* mRNA in *C. elegans*. However, the mechanism for which these short, non-coding RNA molecules lead to gene repression was not understood until Fire and Mello's discovery in 1998.

Ambros and co-workers identified two RNA molecules of different length, one consisting of 61 nucleotides (nt), the other of 22 nt. The longest was predicted to have a stem-loop structure in which the *lin-14* antisense sequence was located within the stem, probably leaving it unable to base-pair with the mRNA. They also found several repeated sequences in the 3’ un-

translated region (3'UTR) of *lin-14* mRNA that were complementary to the 5' end of the shortest *lin-4* RNA, suggesting post-transcriptional regulation of gene expression by antisense RNA binding[63]. It took seven years from this discovery to the identification of the second microRNA, named let-7, in year 2000[64]. Today, 1048 microRNAs have been identified in human and 720 in mouse (*Mus musculus*) (miRBase release 16, September 2010, http://www.mirbase.org)[65].

1.4.1.Transcription of miRNA genes

miRNA genes are located throughout the genome in either intergenic regions or in introns of protein coding genes. miRNAs located in introns of protein coding genes can be transcribed together with the mRNA, and may thus underlie the transcriptional regulation of the corresponding mRNA, or, less common, they can be transcribed by intron-resident promoters independent of host gene promoters[66]. miRNA genes located in intergenic regions are transcribed from their own transcription start site (TSS) and are transcriptionally regulated by their own promoter regions. These miRNAs are transcribed into primary precursor miRNAs (pri-miRNA) that are more than one kilo base-pair in length. Some miRNA genes are located in clusters of 2-7 miRNA genes separated only by a few base-pairs[67]. These miRNAs are transcriptionally regulated by a single promoter, and thus transcribed together as a single, polycistronic pri-miRNA molecule.

The primary transcripts are processed into mature miRNAs in two sequential processing steps, the first taking place in the nucleus and the second in the cytoplasm[68]. miRNA genes are transcribed as capped and polyadenylated[69] transcripts of more than 1 kb in length by RNA polymerase II[70]. Most, if not all, miRNA genes contain the same type of promoters as protein coding genes[71], raising the possibility that microRNA genes can be regulated by the same type of transcription factors as mRNAs, such as enhancers or hormones[72,73]. To understand how miRNAs are involved in regulating different biological pathways, understanding how miRNAs themselves are regulated is important. Investigating effects of known transcription factor's on the expression of miRNAs can provide key information of the complexity of miRNA conducted post-transcriptional regulation.

1.4.2.Maturation of miRNA

Primary miRNA transcripts, pri-miRNAs, are processed into mature microRNAs in two separate steps, both catalysed by RNase III enzymes in complex with dsRNA-binding proteins (Figure 1.5). The first step occurs in the

nucleus, where the pri-miRNA is cleaved by the RNase III enzyme Drosha and its partner DiGeorge syndrome critical region gene 8 (DGCR8), creating a ~70 nt long precursor microRNA (pre-miRNA)[74,75]. Some spliced out mRNA introns in *C. elegans*, *Drosophila melanogaster*, and mammals corresponds precisely to pre-miRNAs. These "mirtrons" circumvent the need for Drosha-DGCR8 complex[75].

The pre-miRNA is exported to the cytoplasm by Exportin-5 (XPO5)[76]. In the cytoplasm the pre-miRNA is further processed into a ~22 nt long duplex (miRNA:miRNA*) by the RNase III enzyme Dicer in complex with TAR RNA binding protein (TRBP)[77,78]. This step can also occur independently of Dicer, but although found throughout the animal kingdom, this is less common[79,62]. One of the strands in this duplex is incorporated into the RNA induced silencing complex (RISC), forming miRISC, and guides sequence specific nuclease activity resulting in degradation of mRNAs, or translational inhibition of mRNAs containing sequences complementary to the miRNA in their 3' untranslated region[80] (3'UTR).

1.4.3.RISC mediated transcriptional regulation

The mature miRNA is loaded into the Argonaut (AGO) effector protein of the RISC assembly. Only one of the strands, the guide strand, directs the silencing complex to mRNAs containing complementary sequences by Watson-Crick base pairing[62]. Generally, the strand with the 5' terminus located at the thermodynamically less stable end of the miRNA:miRNA* duplex is selected to function as the mature miRNA, and the other strand is degraded[75].

The RISC executes post-transcriptional silencing either by sequence specific degradation of mRNA target, or by repression of translation. If there is complete complementarity between the guiding small, non-coding RNA molecule and the targeted mRNA, AGO slices the message at the position opposite the phosphate between the tenth and the ninth base from the 5' end of the guide. This slicing mechanism is the pathway most often used by short interfering RNAs (siRNAs) that is utilised in gene-knockdown technology.

In the case of microRNAs, however, there is not full complementarity between the guide strand and the target strand. The 5'-region has been shown to be the most evolutionary conserved part of the miRNA, suggesting that this region is important in mediating the effect of miRNAs. Indeed, nucleotide 2-8 in the 5'-region of the miRNA has been identified as the most important in prediction of mRNA targets, and these nucleotide region have

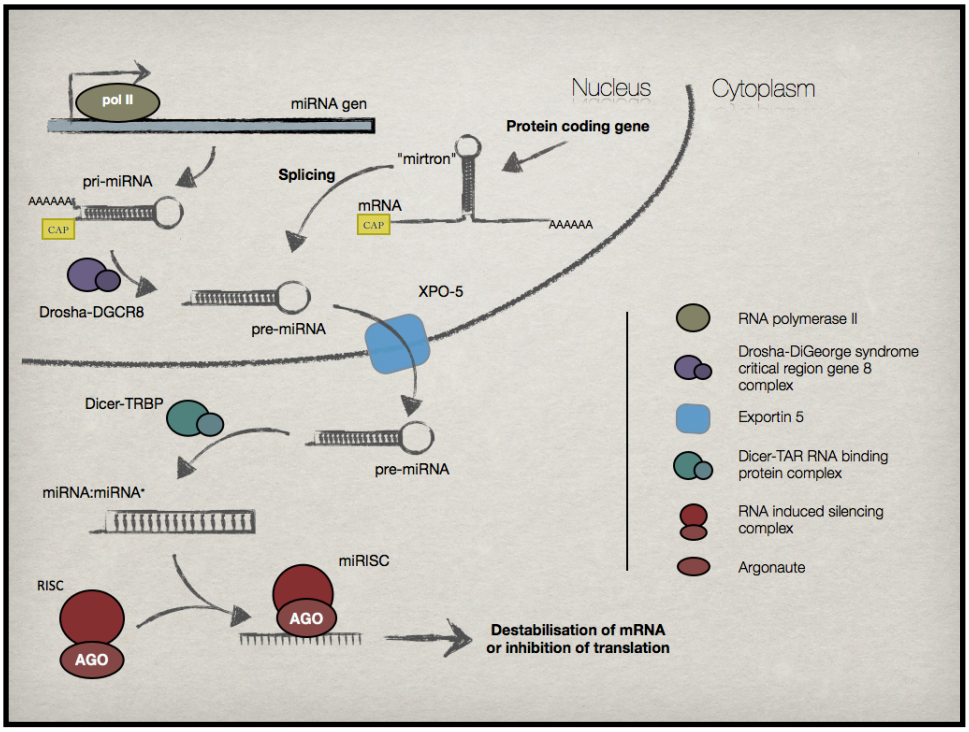


Figure 1.5: MicroRNAs are maturated in two steps executed by different RNase III family enzymes. The mature miRNA:miRNA* duplex is loaded into the RISC, where only one of the strands is used to guide silencing of target mRNAs. The other strand is discarded. Modified from Filipowicz *et al.* (2008)[72].

therefore been named the "seed" region[81]. In the cases where the guiding RNA is not fully complementary with the target, RISC binds and sequesters the transcripts away from the translational machinery into cytoplasmic foci termed P-bodies[82].

Although miRNAs may decrease protein output by direct inhibition of translation, Guo *et al*. (2010) showed that at least 84% of the effect of miRNAs is due to reduction in mRNA levels[83]. The exact mechanism of translational repression is not completely known, but it most likely involves miRNA mediated mRNA deadenylation, which is the initiation step of mRNA catabolism[82,75].

1.4.4. Biological impact of miRNAs

Mammalian miRNAs are predicted to control the expression of more than 60% of all protein coding genes[84]. MicroRNAs seem to have a modest effect of less than 30% down-regulation on most of its targets, although some mRNAs can be down-regulated 50%-80%[85]. However, this does not mean that modest down-

regulation of the targets is out of biological importance. On the contrary the miRNA may be evolutionary conserved due to its effect on dampening and fine tuning, rather than switching off, its target messages. Moreover, since each miRNA can target several mRNAs, and conversely, each mRNA can be targeted by several miRNAs, the cumulative effect miRNA mediated suppression could be of biological significance, even when the effect of the single miRNA is modest[86].

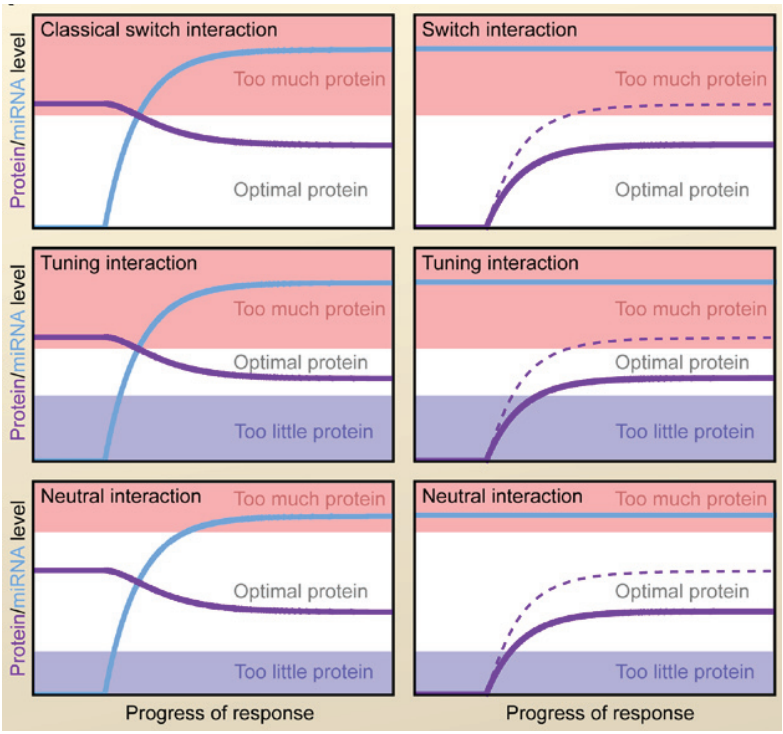
Bartel (2009) has suggested a model of three types of miRNA mediated expression regulation (Figure 1.6). First, miRNA can function in a classical switch, down regulating expression of its target genes to an inconsequent level. This type of regulation occurs for example as to repress aberrant expression of certain genes in certain tissues. Second, as already mentioned, miRNAs may act as to tune, rather than as a binary switch, to dampen protein output to a more optimal level but one that is still functional. Third, neutral interactions may dampen protein output, but this repression is tolerated or offset by feedback mechanisms such that the regulatory sites are under no selective pressure to be retained or lost during evolution. Neutral interactions comprise cases in which biological targeting has no biological function[87].

Even with a down-regulating effect of 30%, miRNAs can have a big impact on mRNA half-life, resulting in a rapid decrease of mRNA[87]. This is of biological importance because it makes cells amenable to quick changes, for instance in response to environmental changes.

1.4.5. miRNA in cancer

Since miRNAs are believed to be involved in almost all biological processes, it is also likely that they are also involved in diseases such as cancer. The expression of many genes have been shown to differ between individuals, and these differences in expression levels have been found to have an inherited component. Since miRNAs targets a number of functionally important protein coding genes, variations in the expression and/or sequence of miRNAs inherited in the germ-line, could be important in cancer predisposition by affecting the expression levels of tumour-suppressor genes or oncogenes[88].

The expression level of miRNAs is found to be globally down regulated in tumour cells, and this has been associated with formation of more aggressive tumours in animals, suggesting that overall loss of miRNA expression enhances tumourigenesis. This global down regulation may function as to repress the expression of miRNAs functioning as tumour suppressors[89].



Figur 1.6: Left panels: in response to a developmental or environmental cue, cells induce miRNA expression (blue), which in turn dampens protein production (purple) from the targeted message. Right panels: protein production from analogous targets also can be dampened by preexisting miRNAs; accumulation in the absence of the miRNA is indicated (dashed line). All panels depict 33% re- pression, which illustrates that classifcation does not rely on the magnitude of repression but instead depends on the properties of each target, namely its threshold(s) for optimal protein output. These thresholds are depicted as the boundaries between the clear (optimal) and colored (suboptimal) regions of the graphs. From Bartel (2009)[84].

A few miRNAs have been shown to be up regulated in tumour cells, functioning as oncogenes or "oncomirs". For example, Martello *et al.* (2010) showed that miR-103/107 expression is induced in breast cancer cells. These miRNAs targets and down regulate Dicer, thus lowering the rate of miRNA maturation. Silencing of miR-103/107 inhibited metastasis, suggesting that a lower miRNA level is necessary for tumour cells to gain metastatic potential[90]. In addition, down regulation of Dicer is associated with poor prognostic impact on survival rate in lung cancer[91].

Another example of oncomirs is the *mir-17~92* cluster, which consists of six miRNAs. This cluster is located in a chromosomal region that is frequently amplified in a subset of human B cell lymphomas, and overexpressed in a variety of human cancers[92].

miRNAs may be used as diagnostic markers for early detection of certain cancer types. In addition, miRNA expression patterns may be a valuable tool in distinguishing histological types of cancer. For example, different miRNA expression patterns have been identified, that may be used to distinguish small cell lung cancer from normal human bronchial epithelial cells and from non-small cell lung cancer[93].

miRNA and the environment

miRNA transcription seems to be tightly regulated in a temporal or tissue-specific manner, and an interesting question is if this regulation can be modulated by environmental factors. The role of miRNAs in complex networks in the field of toxicology is currently largely unknown[84]. Nevertheless, an emerging number of studies are indicating connections between environmental exposure, aberrant miRNA expression, and tumour formation.

In a global miRNA expression analysis of healthy current smokers and never smokers, many miRNAs were found to be down-regulated and their respective mRNA targets were found to be up-regulated[94]. This down-regulation of miRNAs is interesting when compared to the global down-regulation of miRNAs in cancerous tissue, because it indicates that aberrant miRNA expression might be detected, because of exposure, in healthy tissue before tumour formation. This finding suggests that aberrant miRNA expression may be an underlying factor in disease initiation[84].

Izzoti *et al*. (2009) found 126 miRNAs to be down regulated more than a twofold, and only one miRNA to be dramatically up-regulated, in lung in response to exposure of environmental cigarette smoke (ECS). The up-regulated miRNA, miR-294, causes silencing of some transcriptional repressors, thereby resulting in a global increase of gene transcription as response to ECS[95]. Exposure to dioxin was found to result in small changes of miRNA expression in liver, and the authors suggested that hepatic miRNAs are refractory to dioxin[96]. However, the biological effect of small changes in miRNA levels can be of significant magnitude due to the number of potential targets regulated by individual miRNAs[97].

1.5. Aim of study

Little is known about the role of the AHR in regulation of miRNA expression. AHR is a ligand activated transcription factor involved in xenobiotic metabolism, and hence important in mediating cellular response to environmental pro-carcinogens. miRNAs are predicted to be involved in most, if not all, biological processes in the cell,

and may play an important role in this response as negative regulators of gene expression of specific target genes. miRNAs are known to be deregulated in several types of cancer, including cancers of the lung, and knowledge of how miRNA expression is controlled can elucidate their role in tumourigenesis.

The aim of this thesis was to identify the effect of the AHR on miRNA expression. Specifically we wanted to:

- Identify differentially expressed miRNAs in lung of *Ahr* KO and wild type mice
- Study if expression of these also is controlled by the AHR in human lung epithelial cells *in vitro*
- Verify if known *Ahr* regulated miRNAs in murine cells show similar regulation in human lung epithelial cells *in vitro.*
- Explore if miRNAs are subjected to regulation by exposure to chemical carcinogens (B[*a*]P and CSC) in human lung epithelial cells *in vitro*

To identify miRNAs regulated by AHR, an *Ahr* knockout mice experiment was carried out. Expression levels of 750 miRNAs were measured by RT-qPCR array and miRNAs differentially expressed between $Ahr^{-/-}$ and $Ahr^{+/+}$ mice were identified by statistical testing. Bioinformatic analysis was performed to identify predicted targets of the differentially expressed miRNAs. Expression of four selected differentially expressed miRNAs from the mouse experiment, together with three miRNAs previously reported to be differentially expressed in an AHR dependent manner, were further analysed in a RNA interference experiment in human bronchial epithelial cells (HBEC3-KT) *in vitro*.

2. Methods

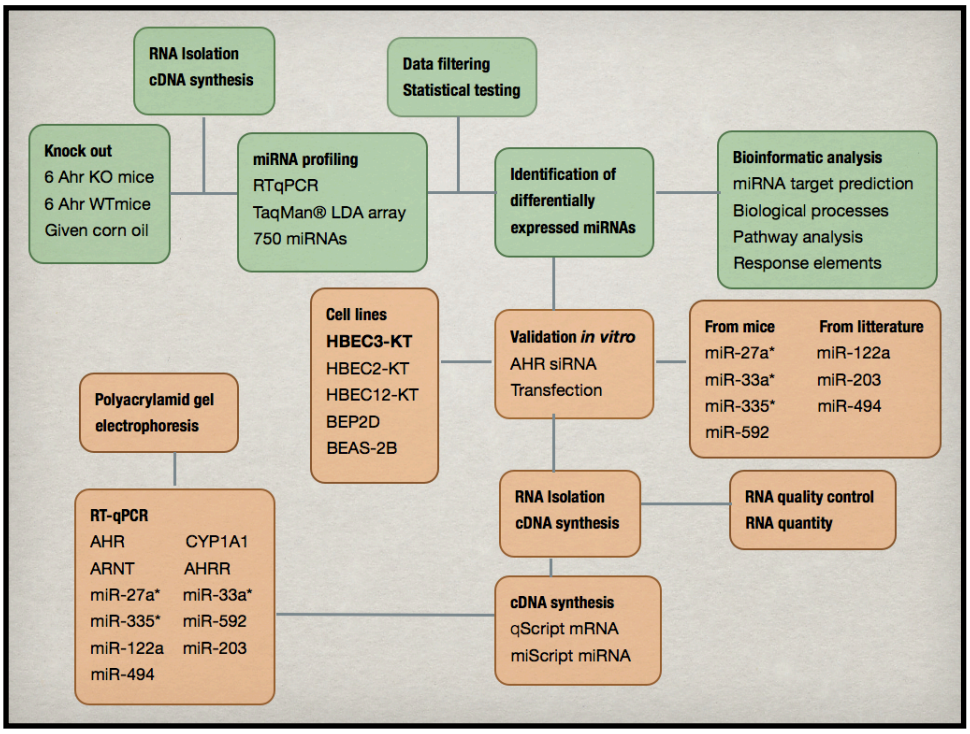


Figure 2.1: Flow chart describing the work flow for identification of AHR regulated miRNAs *in vivo* and *in vitro*.

List of all materials used in the experiments are provided in Appendix A. Figure 2.1 shows the work flow for the experiments performed for this thesis.

2.1.AHR KO Mice Experiment

For analysis of the effect of aryl hydrocarbon receptor (Ahr) on miRNA in mouse lung, expression profiling was performed on an *Ahr* knockout mouse model. The *Ahr* heterozygote model (C57BL6) was interbred to generate

Ahr$^{+/+}$ (wild-type) and *Ahr*$^{-/-}$ (knockout) mice. Genotyping was performed as described in[98]. The mice were maintained at the animal department at STAMI in a germ-free facility using an air-filtered controlled environment. The animals were fed a standard diet (B&K Universal A/S, Norway) and provided water ad libitum.

The mice used in this study were part of the study described in[98], were the effect of B[*a*]P on adduct formation in *Ahr*$^{-/-}$ mice was investigated. The mice were exposed to B[*a*]P solubilised in corn oil (10 mg/mL), while the control group were given pure corn oil (vehicle). In this thesis only the control animals were used. The study consisted of six *Ahr*$^{-/-}$ mice (three male and three female) and six *Ahr*$^{+/+}$ mice (three male and three female, total of 12 mice). The experiment was conducted according to the standards for care and use of experimental animals, and was approved by the Norwegian Reference Centre for Laboratory Animal Science and Alternatives.

Twenty-four hours after oral dose of corn oil the animals were sacrificed. The lungs (as well as other organs such as liver, spleen, kidney, and heart) were harvested, snap frozen in liquid nitrogen, and stored at -80°C.

2.2.Effect of AHR on microRNA expression *in vitro*

To validate results from the *Ahr* knockout mice experiments, expression levels were investigated *in vitro* for the selected miRNAs along with three miRNAs chosen based on their previously reported AHR dependent regulation. These were miR-494, reported to be induced following B[*a*]P exposure in primary murine bronchial epithelial cells *in vitro*[99]; miR-122a has been reported to be induced by TCDD exposure in Hepa-1 mouse cells *in vitro*; and miR-203, reported to have increased expression in Hepa-1 rat cells *in vitro* after TCDD exposure[96].

miRNA expression levels were investigated *in vitro* in immortalised human bronchial epithelial cell line (HBEC3-KT). This cell line is established from a 65 year old female smoker with no cancer, immortalised by overexpressing two genes associated with growth arrest. Telomerase reverse transcriptase (*TERT*) is overexpressed to prevent telomere shortening and bypass telomere-dependent senescence, and cyclin dependent kinase 4 (*CDK4*) is overexpressed to bypass premature growth arrest mediated by CDK inhibitor p16INK4a[100].

Three independent experiments were conducted, each containing two replicates of each sample and control.

Samples were transfected with AHR short-interfering RNA (siRNA) in order to knock down AHR gene function. Control cells were transfected with two control siRNA molecules with different GC level. These controls are of the same length as the gene specific siRNAs, but they do not have any known targets in the human genome. Cells were subsequently exposed to B[*a*]P, CSC or dimethyl sulfoxid (DMSO, vehicle) for one day or 5 days. Total RNA was isolated using TRIzol-RNA lysis reagent, and qScript (Quanta) and miScript (Qiagen) was used for RT-qPCR analysis of mRNA and microRNA, respectively (see sections below).

2.3.Cell culture

See appendix B for information on general cell culture work.

Human bronchial epithelial cells were taken from STAMIs cell bank, were they were stored in liquid nitrogen. All cells were cultured in LHC-9 medium (Invitrogen) added 1% penicillin and streptomycin, unless stated otherwise. All cells were grown on collagen coated dishes (see appendix B.3). Incubation conditions were 37°C and 5% CO_2. All cell culture work was performed under sterile conditions in OAS LAF benches.

2.3.1. Seeding of cells for experiment

Cells seeded for transfection experiment were cultured in LHC-9 medium without penicillin and streptomycin, because antibiotics may cause undesired toxicity during transfection.

1. The media over the cells was removed.
2. The petridish was washed twice with PBS (2 x 10 ml).
3. Trypsin (1 ml, Invitrogen) was added to the dish, and the cells were incubated (37°C) until they had loosened from the dish. This was checked in the microscope.
4. LHC-9 medium (5 ml) was added to the dish and the cell suspension was pipetted to a centrifuge tube and centrifuged (1000 rpm, 4 minutes).
5. The supernatant was removed.
6. The pellet was re-suspended in LHC-9 without pen./strep. Amount of medium used depended on the desired dilution of the cells.
7. Cells were counted using automated cell counter (Countess, Invitrogen).
8. $2.5x10^5$ cells were seeded in 1 ml LHC-9 in each 35 mm well of a six-well plate.
9. Cells were incubated for 24 hours.

2.3.2. siRNA mediated AHR knock-down

siRNAs are double stranded RNA molecules of 21 nucleotides that are widely used for gene specific tran-

script knockdown. After introduction into cells, they are processed by the cells own RNA interference machinery. The dsRNA molecules are usually introduced into the cells by either lipid-mediated transfection or viral-mediated transduction (there are other methods as well, but these are the two most common). Which one of these methods that is the best one to use depends on the cell type being studied. In this study, a cationic lipid based reagent was used as transfection method for the delivery of AHR specific siRNAs because of its suitability for delivering molecules across a diverse range of commonly used cell types. AHR siRNA sequences are shown in Table 2.1.

1. Stealth Select RNAi AHR siRNA 2.49 µl per well in a 6-well plate) was mixed with Opti-MEM® I Reduced Serum Medium (247.5 µL/well).
2. Lipofectamine™ RNAiMAX (4.8 µL/well) was diluted with Opti-MEM® I Reduced Serum Medium (245.2 µL/well).
3. The diluted siRNA was added to the diluted Lipofectamine RNAiMAX and mixed gently and incubated for 20 minutes.
4. Medium was removed from the wells and DMEM (2 mL) was added.
5. The transfection solution was carefully dripped over the cells (500 µL/well) and the plates were gently rugged. The final conentration of siRNA was 20 nM.
6. Cells were incubated for 48 hours (37°C, 5% CO2)
7. The transfection solution was removed and the cells were either exposed to B[a]P, CSC, or DMSO for 1, 3, or 5 days.
8. Cell culture experiments were stopped by placing the multi-dishes on ice, removing the medium over the cells, followed by washing with PBS (3 x 2 mL). The multi-dishes were stored at -80°C until RNA-isolation.

2.3.3.The persistence of siRNA mediated gene silencing.

To examine the durability of siRNA mediated AHR silencing, cells where transfected for 48 hours as described above, and then stopped im-

Table 2.1: Sequences of control and AHR specific siRNA used in AHR knock down experiment in HBEC3-KT *in vitro.* All siRNA molecules were suspended in 1 mL RNase-free H2O, Tris-HCl (10 mM, pH8.0), NaCl (20 mM), EDTA (1 mM).

	Sequence
AHR fwd 1	CCUUUAAUGGAGAGGUGCUUCAUAU
AHR reverse1	AUAUGAAGCACCUCUCCAUUAAAGG
AHR fwd 2	GAGGCUCAGGUUAUCAGUUUAUUCA
AHR rev 2	UGAAUAAACUGAUAACCUGAGCCUC
AHR fwd 3	CCAAGCGGCAUAGAGACCGACUUAA
AHR rev 3	UUAAGUCGGUCUCUAUGCCGCUUGG

mediately, 1 day, 3 days, or 5 days after transfection solution was removed.

2.3.4.Toxicity testing

The CellTiter-Blue® cell viability assay (Promega) uses the indicator dye resazurin to measure the metabolic capacity of cells. Viable cells retain the ability to reduce resazurin into resorufin, which is highly fluorescent. Nonviable cells rapidly lose metabolic capacity, do not reduce the indicator dye, and thus do not generate a fluorescent signal.

1. Cells were passaged to two 96-well plates ($1.1x10^4$ cells/well) and incubated (37°C, 5% CO2) for 24 h before transfected as described earlier (250 µl transfection media).
2. When the transfection medium was removed, one plate was added LHC-9 medium and incubated for another 24 h before toxicity testing. The second 96-well plate was added LHC-9 (100 µL) and CellTiter-Blue® (20 µl) and icubated (4 h).
3. Fluoresence was measured at 560/590nm (Modulus microplate reader, Pronova).

2.4.RNA-isolation

RNA from both harvested mouse lung and from human lung cell cultures was isolated using TRIzol-RNA reagent, a monophasic solution of phenol and guanidine thiocyanate that disrupt and homogenise cells and tissues and inhibit RNases. TRIzol-RNA isolates total RNA and is thus usable for isolation of both mRNA and miRNA. Chloroform is added to the TRIzol-RNA homogenate and the solution is centrifuged, which leads to the separation of three phases. The aqueous (upper) phase contains RNA, the interphase contains DNA, and the organic (lower) phase contains proteins. The RNA is precipitated with isopropanol

Protocol

1. The 6-well multi-dishes were taken from the freezer and placed on ice. Isol-RNA Lysis Reagent (1 mL/35 mm well) was added.
2. A cell scraper was used to loosen the cells from the plate. The lysate from each well was transferred to a twist top vial and incubated (room temperature, 5 min).
3. Chloroform (0.2 mL) was added to each vial, and the vials were shaken by hand for 15 seconds before incubated (room temperature, 2-3 minutes).
4. The vials were centrifuged (12000 g, 4°C, 15 min).
5. The upper, aqueous phase was transferred to a new twist top vial, added iso-propanol (0.5 mL) and incubated (room temperature, 10 min). The lower, organic phase was discarded.

6. The vials were centrifuged (12000 g, 4°C, 15 min).
7. The supernatant was aspirated and discarded.
8. The pellet was washed with ethanol (75%, 1mL) and vortexed.
9. The vials were centrifuged (12000g, 4°C, 5 min).
10. The supernatant was discarded, and the pellet was dried (15-20 min) in the vial without the cap on.
11. The pellet was re-suspended in nuclease free water (10-15 µL, depending on the pellet size) and put on ice.
12. The vials were incubated (65°C, 10 min), spun down, mixed and put on ice.
13. The vials were put in freezer (-80°C) for minimum one night.
14. The RNA-samples were incubated (65°C, 10 min).
15. RNA was stored at -80°C.

RNA consentration

RNA-concentration was determined by OD (260 nm/280 nm) in an Eppendorf Biophotometer. Knowing the concentration of RNA is important to obtain correct dilution in further work. Nucleic acids have absorption maximua at 260 nm. A 260 absobtion of 1.0 is equivalent to 40.0 µg/ml RNA (and 50.0 µg/ml dsDNA). Pure RNA has a 260/280 ratio of 2.0, while pure DNA has a ratio of ~1.8. Each sample was diluted 70x (1 µL RNA sample + 69 µL TE-buffer) before OD measurement in a quartz cuvette.

RNA quality

RNA quality was measured on a subset of RNA samples by using Aligent Bioanalyzer RNA 6000 Nano Kit (No. 5067-1511). Results are shown in Appendix E.

2.5. RT-qPCR

The polymerase Chain Reaction (PCR) is a method for amplification of DNA-sequences using repeated cycles of altering high and low temperatures[101]. Reverse transcriptase quantitative PCR (RT-qPCR) is a much-used method for examining gene expression. RT-qPCR uses a fluorescent reporter molecule that binds the PCR product and reports its presence by fluorescence. As the amount of product increases exponentially during the PCR, the fluorescent signal also increases exponentially, reflecting the amount of product produced[102]. In the work for this thesis two different fluorescent reporters were used; Hydrolisation probes (TaqMan LDA cards, Applied Biosystems) for detection of miRNAs in mouse lung, and SYBR Green for detection of mRNA and miRNA in immortalised human lung cell lines.

2.5.1. cDNA synthesis of mRNA

Complementary DNA is synthesised from RNA by a Reverse Transcriptase (RT) enzyme. For cDNA synthesis qScript cDNA Synthesis Kit (Quanta) was used. This kit contains a combination of oligo-dT and random primers, a MMLV RT enzyme, and a ribonuclease inhibitor protein.

Protocol

1. A master mix was made of qScript RT-enzyme (1µL per sample), qScript cDNA SuperMix (4 µL per sample), and DEPC water (14.0 µL pr. sample). An eppendorf tube for each RNA-sample was added master mix (19 µL) and RNA (1µL). The tubes were vortexed.
2. The eppendorf tubes were incubated in the thermal cycler at the following program:
    - Incubation, 22°C, 5 minutes.
    - Synthesis step, 42°C, 30 minutes.
    - Incubation, 85°C, 5 minutes.
    - Incubation, 4°C, minimum 5 minutes.
3. After PCR each tube was added TE-buffer (80 µL). The cDNA was stored at -20°C.

2.5.2. cDNA synthesis of miRNA

For cDNA synthesis of miRNAs, miScript kit (Qiagen) was used. Mature miRNAs are not poly-adenylated, and consequently oligo-dT primers cannot be used directly for cDNA synthesis. Therefore, poly(A) polymerase is added to the mix to synthesize a poly(A) tail on all non-coding RNAs. Oligo-dT and random primers can then be used for cDNA synthesis of non-coding RNAs as well as mRNAs. The polyadenylation and reverse transcription is performed in a single step. The oligo-dT primers in the miScript kit have a universal tag sequence on the 5' end, which is utilized in qPCR of miRNAs (section 2.5.4).

Protocol

1. RNA templates were thawed on ice. miScript RT buffer and RNase free water were thawed at room temperature
2. The RNA of the biological replicates in each experiment was mixed before cDNA synthesis.
3. A master mix was made, containing (volume/reaction):
4. miScript RT Buffer, 5X (4 µl)
5. RNase-free-water (14.5 µl)
6. miScript Revese transcriptase (1 µl)
7. Master mix (20 µl) was added to empty plastic tubes
8. RNA (0.5 µl) was added to the tubes with master mix, vortexed and put on ice.
9. The tubes were incubated in a cycler (37°C, 60 min) and then incubated in a heating block (95°C, 5 min) to inactivate the enzymes.

10.cDNA was diluted with TE-buffer (30 µl) and then stored at -20°C until qPCR.

2.5.3. mRNA qPCR

qPCR of mRNA is performed by using two primers in opposite directions of the desired gene. The primers are designed to span introns, such that amplification of genomic DNA residues in the cDNA sample is avoided.

Protocol

To obtain optimal Cq values between 20-30 cycles for all genes, the cDNA was diluted to appropriate concentrations for each gene. β-actin was used as a reference gene for normalisation of gene expression. This gene in stably and highly expressed in most cells, and therefore an cDNA amount equivalent to 5 ng RNA input in the cDNA synthesis, was used in the qPCR reaction. The amount of *CYP1A1* cDNA was 100 ng since this gene has a low expression. For all other genes 50 ng were used. Concentration of all primers was 25 pmol/µl.

For mRNA qPCR PerfeCTa SYBR Green FastMix, ROX (Quanta) was used. This kit contains AccuFast™ Taq DNA polymerase, SYBR Green I dye, ROX as an internal fluorescence reference, along with buffer and stabilisers.

qPCR of mRNA was setup manually in 96-well plates. A master mix of 36 µl per sample was made:

- dH_2O (12.08 µl)
- 2x PerfeCTa SYBR Green (23.0 µl)
- Primer up (25 pmol/µl, 0.46 µl)
- Primer lo (25 pmol/µl, 0.46 µl)

A 4x dilution sereis was used to generate a standard curve. In PCR tube strips, 36 µl of master mix was added cDNA (10 µl), dH_2O (10 µl) for non-template control (NTC) or cDNA from 4x dilution series (10 µl). All samples were run in 2 parallelles of 20 µl reactions on a 96-well plate. Cycling conditions are shown in table 2.1.

Table 2.1: Cycling conditions for qPCR with PerfeCTa kit (Quanta).

Step	Temperature (°C)	Time	
AccuFast Taq polymerase activation	95	2 min.	
Denaturation of cDNA	95	10 sec.	40 cycles
Primer annealing and extension	60	30 sec.	

Melting curve analysis was performed at the end of each qPCR to control for correct PCR product.

2.5.4.miRNA qPCR.

RT-qPCR detection of miRNAs is performed by using two primers, one that is miRNA specific (this is normally the entire mature miRNA sequence) and the other is a universal primer that is complementary to the tag sequence added in the cDNA synthesis step.

RNU6B was used as reference gene in normalisation of relative miRNA expression.

Protocol miScript

qPCR of miRNAs was setup using pipetting robot (EP-Motion, Eppendorf). qPCR was performed in 384-well plates with reaction volumes of 12 µl:

- 2x QuantiTect SYBR Green PCR Mix (6 µl)
- 10x miScript Universal primer (1.1 µl)
- 10x miScript Primer assay (miRNA specific, 1.1 µl)
- RNase-free water [1.3 µl)
- Template cDNA (2.5 µl)

Cycling conditions are shown in table 2.2. Melting curve analysis was performed at the end of each qPCR.

Table 2.2: Cycling conditions for miScript qPCR assay (Qiagen).

Step	Temperature (°C)	Time	
HotStart Taq Polymerase activation	95	15 min.	
Denaturation of cDNA	95	10 sec.	40 cycles
Primer annealing	94	30 sec.	
Extension	70	30 sec.	

Normalisation of expression data from cell line experiments

Gene expression was quantified by calculating relative changes in mRNA or miRNA level between samples. The quantification cycle (C_q) value is the cycle number at which the fluorescence in a sample reaches a manually set threshold. The C_q value for each gene of interest (GOI) is normalised with the C_q value of a reference gene.

The C_q determination of the threshold level and in consequence the accuracy of the quantification results are influenced by the amplification efficiency. Changes in the PCR efficiencies are caused by RT and PCR inhibitors or enhancers, and by variations in the RNA pattern extracted. Even small efficiency differences between GOI and reference gene generate false expression ratio[103]. Therefore, an efficiency corrected mathematical model was used for calculation of relative expression for mRNA (Equation 1)[104].

$$\text{Relative expression} = \frac{(E_{ref})^{C_{q,ref}}}{(E_{\text{target}})^{C_{q,\text{target}}}} \quad (1)$$

Here E is the PCR efficiency, ref is the reference gene, and target is GOI. The efficiency of the PCR is found by including a dilution series for generating a standard curve in every qPCR. The slope of the standard curve is used to calculate the efficiency in Equation 2[103].

$$E = 10^{-1/slope} \quad (2)$$

If the amount of PCR product doubles at every cycle (E=2) the slope of the standard curve will be -3.33. For miRNA qPCR no standard curve was included and 100% efficiency (E=2) was assumed. Fold change was calculated by dividing relative expression values of each sample (B[*a*]P or CSC exposed) on the expression level value of the calibrator (DMSO exposed).

2.5.5.MicroRNA analysis on TaqMan® Low-Density Arrays (LDA)

Global mouse miRNA expression profiling was conducted using two TaqMan Low-Density (LDA) array cards (Applied Biosystems, card A+B version 3, 4444913). These two LDA cards cover almost all known mouse miRNAs of Sanger miRBase build 15 [65] Each TaqMan LDA card consists of 384 wells containing 375 preloaded mouse miRNA targets and six controls – the expression of 750 mouse miRNA can be quantified simultaneously in the two LDA cards. The preloaded TaqMan Array is processed in two-step RT-qPCR. During the first step, total RNA (containing small RNA species) is reverse-transcribed followed by RT-qPCR for detection of specific cDNA sequence.

The total RNA isolated from the six *Ahr*$^{-/-}$ mice (three male and three female) and six *Ahr*$^{+/+}$ mice (three male and three female, in total 12 mice, see section 2.1) were run in the LDA card sets (A+B). The LDA expression measurements were carried out in collaboration with Dr. Mogens Kruhøffer, CSO, at AROS Applied Biotechnology AS, Science Park Skejby, Aarhus, Denmark. The raw miRNA expression data (the C_q-values) generated for each LDA card were further processed at STAMI in collaboration with Dr. Nur Duale at the Norwegian Institute of Public Health.

Normalisation of expression data

High throughput quantitative reverse transcriptase PCR (RT-qPCR) is becoming a widely used method for expression profiling of microRNAs[105]. In such experiments, there are several variables that needs to be accounted for, such as differences in amount of starting material (input cDNA), differences in enzymatic efficiencies, or differences of expression between tissues or cells, when conducting a gene expression experiments[106]. The variation in gene expression data is a combination of biological and technical variations, and to reduce the influence of the technical differences the gene expression data should be normalised towards one or more references, such as reference genes that ideally do not vary between the different samples in the individual experiment. The purpose of normalisation is to reduce the technical variation within a dataset, ena-

bling a better appreciation of the biological variation[107].

There are several different ways to normalise high-throughput RT-qPCR microRNA data, and no one method has so far become standard. The most used approach is normalisation to a reference gene, the same method as used in relative mRNA expression analysis in RT-qPCR. In miRNA expression analysis, stably expressed miRNAs or other short RNAs, such as short-nucleolar RNA (snoRNA) or short nuclear RNA (snRNA, for example RNU6[105,108]) are often used for normalisation.

No universal stably expressed miRNAs has been established. However, it is common to use the microRNAs that are most stably expressed across the samples in each experiment for normalisation. There are publicly available software's such as GeNorm [106] and NormFinder[109], that can be used to determine stably expressed miRNAs. These algorithms identify the miRNAs that are most stably expressed across all samples in the experiment. In this thesis, we used the GeNorm which is an extension to Excel (Microsoft Office, Microsoft), that identifies the most stably expressed miRNAs and then uses the geometric mean of these to calculate a normalisation factor (NF) for each sample (each mouse in this case). The input values in GeNorm are quantities, calculated by the delta-Cq method where the highest relative expression for each gene is set to 1 (Equation 3).

$$Q = 2^{-(C_{q,i} - C_{q,\min})} \qquad (3)$$

In this equation, Q is the calculated quantity for the gene of interest (GOI), $C_{q,i}$ is the C_q value for sample *i*, and C_q,min is the minimum C_q value for the GOI across all samples. The quantities are then used by the GeNorm application to identify the most stably expressed genes by determining each gene's pairwise variation with all other genes as the standard deviation of the logarithmically transformed expression ratios. By taking the average pairwise variation of a particular gene with all other control genes an internal control gene-stability factor, M, is calculated. The miRNA with the lowest M has the most stable expression across all samples. The genes with the highest M is stepwise excluded resulting in two constitutively expressed reference genes that have the most stable expression in the tested samples. NF is then calculated by taking the geometric mean of three or more of the most stably expressed genes.

The expression level of the GOI is then calculated by dividing the GOI quantity with the normalisation factor (Equation 4).

$$NQ_{GOI} = \frac{Q_{GOI,i}}{NF_i} \qquad (4)$$

where NQ_{GOI} is the normalised quantity value for the gene of interest, $Q_{GOI,i}$ is the quantity of the gene of interest in sample i, and NF_i is the normalisation factor for sample i.

Another approach for normalisation of high-throughput miRNA expression analysis is normalisation against the global mean of microRNA expression level (see equation 5)[107]. This approach, however, assumes that the overall expression level of microRNAs remains constant under the conditions in the experiment.

$$\text{Expression level} = 2^{-(C_{q,GOI\,i} - C_{q,GlobalMean\,i})} \quad (5)$$

One advantage of the "Global mean" method is that all raw Cq–values are normalized to the same value, making it possible to compare expression levels of different miRNAs. In contrast, in the "GeNorm" method the Cq–values are first transformed to quantities, meaning the expression level of each gene is calculated relative to the highest expression level across all samples for that gene. If comparison of expression levels between miRNAs is not of interest, this is an unimportant feature.

Vandesompele and co-workers, who also designed the GeNorm algorithm, first described the "Global Mean" normalization method in 2009[107]. They argued that the global expression level of miRNAs remains constant across all samples, and tested this by running GeNorm and NorfmFinder on the mean. However, the global mean may not remain constant across all samples for all individual experiments, so it may be necessary to evaluate the stability of the mean for each experiment.

In this thesis, normalisation was done by both the global mean method and by the GeNorm method on the mouse miRNA expression data. Statistical analyses were performed on data sets normalised by both these methods and the results were compared to increase the robustness of the conclusions.

In this thesis, normalisation was done by both the global mean method and by the GeNorm method on the mouse microRNA expression data. Statistical analyses were performed on data sets normalised by both these methods and the results were compared to increase the robustness of the conclusions.

2.6.Statistical methods

Statistical analyses were performed using PASW Statistics v.18.0 (SPSS Inc., IL, USA) and $p > 0.05$ was accepted as statistically significant. Watson U^2 test was used to evaluate whether the data showed a normal distribution. Comparison of two groups was performed using parametric two-sample *t*-test. Comparison of multiple

groups was performed using one-way ANOVA followed by a post hoc Bonferroni test to allow for multiple comparisons.

To test for differences between the sample groups, statistical analysis were performed on both the GeNorm normalised data set and the global mean normalised dataset. T-test was performed in PASW Statistics to check for microRNAs differentially expressed between *Ahr*$^{+/+}$ mice and *Ahr*$^{-/-}$ mice independent of sex, and between male and female mice independent of genotype.

Two assumptions for the t-test and ANOVA are that both groups of ΔC_q must have Gaussian distributions and equal variances. However, these assumptions are not valid in many RT-qPCR experiments using realistically small sample sizes. Therefore, as a non-parametric alternative to the unpaired *t*-test, a Mann-Whitney-Wilcoxon test may be a more robust and appropriate alternative to analyse the two pools of ΔC_q-values. On the other hand, gene expression data can become closer to normal distribution after logarithmic transformation making a parametric test more appropriate. Moreover, on small data sets, the non-parametric tests lack statistical power[110].

ANOVA analyses the variance within and between the groups and returns an F-ratio that can be used to define a p-value. If the p-value is low ($< \alpha = 0.05$), the null-hypothesis that the groups have the same mean is rejected. While *t*-test and Mann-Whitney-Wilcoxon test can be used for hypothesis testing of two groups, the ANOVA is the appropriate test when comparing three or more groups. One advantage of ANOVA is the possibility of post hoc tests, which identifies which group differs from each other. One-way ANOVA was performed on the two normalised datasets with Bonferroni post hoc tests.

2.7.Hierarchical clustering and principal component analysis (PCA)

Hierarchical clustering and principal component analysis (PCA) were performed using JExpress Pro v.2009 (MolMine A/S, Bergen, Norway). PCA reduces the dimensionality and classifies variables and thus helps to identify the variables that contribute most to the variation in a dataset[111]. Hierarchical clustering subdivides the whole group of genes or treatments in such a way that similar groups fall into the same cluster, whereas dissimilar fall in different clusters. This can assist us in identifying genes or treatments with similar response[112]. We used the

Euclidean distance metric and the average linkage cluster method.

To visualise differences in the expression levels between *Ahr*$^{+/+}$ and *Ahr*$^{-/-}$ mice a heatmap was created for the expression values of the significantly differentially expressed miRNAs as identified by *t*-test or ANOVA. In addition, a component variance analysis was performed to identify grouping/clustering of miRNAs and of the mice.

2.8. Analysis of miRNA expression data

2.8.1.Genome database search for AHR response elemnts

The AHR/ARNT complex binds to specific sites in the promoter region of its target genes. There are two identified binding elements, AHR Response Element I (AHRE-I or XRE) and Aryl Hydrocarbon Response Element II (AHRE-II) (see consensus sequences in table Table 2.3). The core binding sequence in XRE, GCGTG, is shown in bold. A search in the human and mouse genomes for AHR/ARNT binding sites in positions -10,000 bp to + 1,000 bp of the differentially expressed miRNAs was performed for both AHRE-I (core and extended sequence) and AHRE-II (Mathematica 8, see Appendix E for search input/search algorithms). No evolutionary conservation analysis was performed.

Table 2.3: AHRE-I and AHRE-II sequence.

AHRE-I	5'-T/GN**GCGTG**A/CG/CA-3’
AHRE-II	5’-CATG-(N6)-CT/ATG-3’

2.9.miRNA target prediction and functional analysis

Predicted targets for the 10 most significantly differentially expressed microRNAs were downloaded from miRWalk database (available at http://www.ma.uni-heidelberg.de/apps/zmf/mirwalk/index.html). In addition 10 non-differentially expressed miRNAs chosen by random from the remaining miRNAs on the LDA cards, were chosen as reference, and predicted targets for these were also downloaded. A pivot table was created in Excel for both the reference gene set and the differentially expressed gene set in order to identify genes targeted by several miRNAs within each set. Genes targeted by more than three miRNAs were used for pathway-enriched analyses (Metacore GeneGo, Thomson-Reuters). The resulting table of pathways for both differentially expressed miRNAs and for the reference miRNAs were imported in Excel, and the ratio between the p-value between DE miRNAs and ref-miRNAs was calculated for each pathway.

3. Results

3.1.MicroRNA expression in *Ahr* wild-type and knockout mice

3.1.1.Pre-processing of miRNA expression raw data

The miRNA expression levels in lungs of *Ahr* wild-type (n=6) and knockout (n=6) mice were analysed by RT-qPCR on LDA cards. Raw C_q–values from 750 miRNAs were pre-processed to remove outliers and miRNAs with inadequate measurements. The C_q values equal to 35 cycles were considered as a limit of detection and all C_q values > 35 were removed from downstream analysis. In addition, filtering criteria for missing values was set to 50%, which is the minimum percentage of existing values, and all the patterns with less than 50% existing values were removed. The outcome of these filtering criteria was a miRNA expression matrix consisting of 565 miRNAs X 12 samples.

From the 565 miRNAs, the 32 miRNAs with lowest variance were selected in order to identify the most stably expressed microRNAs using GeNorm. The stability of these miRNAs across all samples was analysed by calculation of an M-value (see section 2.5.5). The stability of the global mean was also evaluated along with the 32 miRNAs. The expression levels of mammalian RNU6 were found to be the most stable across all samples, and they were found to be even more stable than the global mean (see Appendix C). The GeNorm software calculated a normalisation factor for each sample by taking the geometric mean of the quantities of the most stably expressed microRNAs (see Appendix C). The miRNAs were normalised by dividing the miRNA quantity with the NF generated by GeNorm. Since the stabilities of mammalian RNU6 were higher than the stability of the global mean, GeNorm normalised data were used for downstream analysis.

3.1.2. Identification of differentially expressed miRNAs

Distribution fit tests were performed in Mathematica 8 (Wolfram Research, Inc.) and 24 and 12 miRNAs from the wild type and knockout dataset were identified, respectively, as not normal distributed at the 2.5% level

Table 3.1: miRNAs significantly ($p<0.05$) differentially expressed between *Ahr* wild-type and knockout mice. Fold change was calculated by dividing relative expression levels of knockout on relative expression levels in wild-type.

miRNA Name	T-Test	Fold Change
mmu-miR-335-3p	2.22E-04	0.39
mmu-miR-592	4.67E-04	2.46
mmu-miR-33*	5.60E-03	0.76
mmu-miR-27a*	2.34E-03	0.70
mmu-miR-210	0.020	0.73
mmu-miR-764-5p	0.020	0.76
mmu-miR-491	0.023	0.78
mmu-miR-188-5p	0.027	0.64
mmu-miR-183*	0.017	0.75
mmu-miR-183	0.025	0.76
mmu-miR-15a*	0.028	0.80
mmu-miR-223	0.028	0.76
mmu-miR-148b	0.041	0.74
mmu-miR-182	0.043	0.74
mmu-miR-547	0.046	0.69

(The Mathematica software only performs *t*-test on datasets that are normal distributed with p>0.025 according to the Watson U^2 test). However, due to the vast majority of the miRNAs being normal distributed, *t*-test was chosen to identify if any of the 565 miRNAs showed indications of being differentially expressed between $Ahr^{-/-}$ and $Ahr^{+/+}$.

The microRNAs found to be significantly diffenrently expressed (*t*-test,

Table 3.2: Results from ANOVA testing on miRNA expression levels. A Bonferroni post hoc test was performed to identify which groups significantly differed from the other groups. Fold change was calculated by dividing the relative expression level of knockout on relative expression level of wild-type for each group. Fold change is given only for the groups in which the particular miRNA was found to be differentially expressed. FKO: Female knockout, FWT: Female wild-type, MKO: Male knockout, MWT: Male wild-type.

miRNA Name	p-value	FKO/ FWT	FKO/ MKO	MKO/ MWT	FKO/ MWT	MKO/ FWT	MWT/ FWT	KO/WT
mmu-miR-191*	0.0016	1.42	1.45	1.37	1.06			1.03
mmu-miR-592	0.0029	2.83		2.15	2.46	2.47		2.46
mmu-miR-335-3p	0.0032	0.44		0.34	0.36	0.43		0.39
mmu-miR-141*	0.0058		0.64	0.69				
mmu-miR-154*	0.0107		0.53		2.48			
mmu-miR-676*	0.0117			0.66				
mmu-miR-27a*	0.0148			0.64				
mmu-miR-24	0.0167			0.77				
mmu-miR-183*	0.0218				0.64			
mmu-miR-135a	0.0283			0.46				
mmu-let-7b*	0.0295						1.28	
mmu-miR-764-5p	0.0302					0.73		
mmu-miR-701	0.0304			0.70				
mmu-miR-101b	0.0310			0.63				
mmu-miR-188-5p	0.0337			0.51				
mmu-miR-448	0.0390		0.48					
mmu-let-7g*	0.0447			0.61				

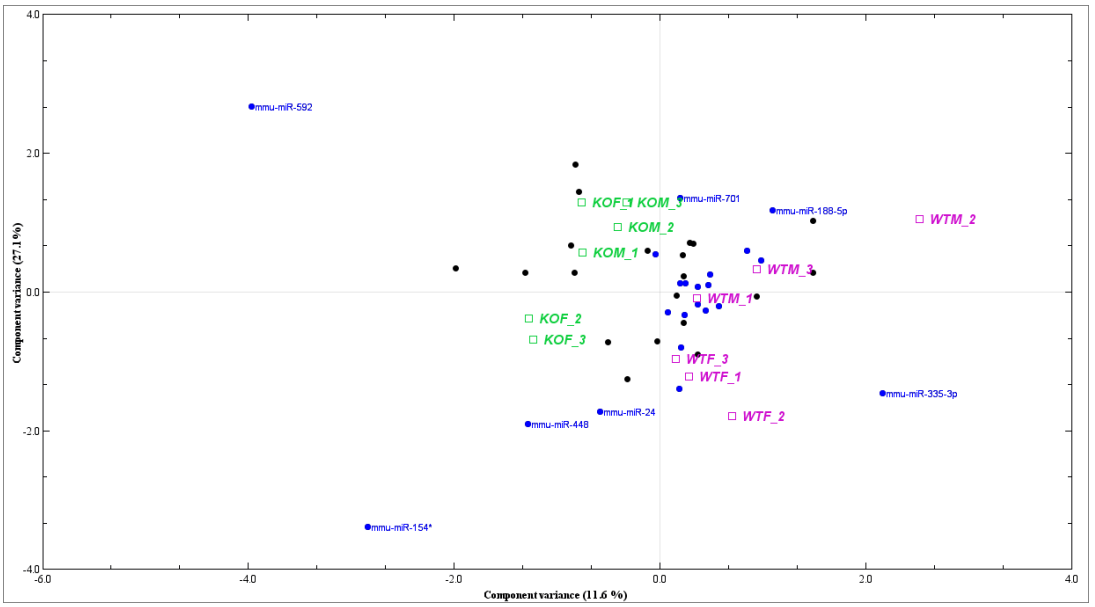


Figure 3.1: Principal component analysis of the significantly differentially expressed microRNAs as defined by t-testing of both GeNorm normalised data and Global mean normalized data sets. The graph shows that the wild type mice (WT) clusters on the right side of the y-axis, while the knock out mice (KO) clusters on the left side of the y-axis. The graph indicates that miR-592 and miR-335-3p are opposite correlated, which agree whit miR-592 being significantly up-regulated, and miR-335-3p being significantly down regulated (Table 3.1). WTF: wild-type female, WTM: wild-type male, KOF: knockout female, KOM: knockout male.

$p<0.05$) are shown in Table 3.1. Fifteen microRNAs were differentially expressed between *Ahr*$^{+/+}$ and *Ahr*$^{-/-}$ mice. Fold change was calculated (see appendix B), and only one miRNA (miR-592) was found to be up regulated in knockout mice, while all others were down regulated. ANOVA test was also performed to identify significant differences between all four groups in the experiment; *Ahr*$^{+/+}$ male and female, and *Ahr*$^{-/-}$ male and female. P-values from ANOVA is shown in Table 3.2. Fold change for the groups found to be significantly different from each other, identified by Bonferroni *post hoc.* is shown in Table 3.2.

3.1.3. Principal component analysis and clustering

Principal component analysis (PCA, Figure 3.1) and Hierarchical clustering (figure 3.2) were performed (J-Express 2009, MolMine A/S, Bergen, Norway) on 42 miRNAs, including those differentially expressed (*t*-test) and the miRNAs from the ANOVA with $p<0.05$. For comparison, the 565 miRNAs were also normalised by using the global mean method, and the same statistical tests was performed on this normalised dataset as for the GeNorm normalised dataset. While most of the

differentially expressed miRNAs found in the global mean dataset overlapped with the differentially expressed in the GeNorm dataset, 10 additional miRNAs were found to be differentially expressed in the global mean dataset. These 10 miRNAs were included in the clustering and principal component analysis, resulting in a total of 42 miRNAs.

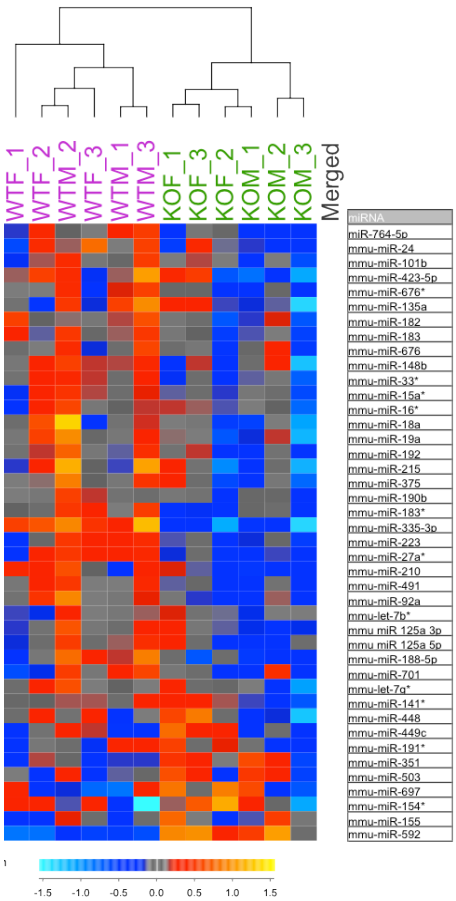


Figure 3.2: Heat-map showing relative expression levels of microRNAs differentially expressed in Ahr+/+ and Ahr-/- mice. Clustering analysis separate knock out from wild type mice, and the figure indicates a tendency towards most of the significantly deregulated microRNAs being down-regulated in Ahr-/- mice. WTF: wild-type female, WTM: wild-type male, KOF: knockout female, KOM: knockout male.

The heat-map created for the 42 miRNAs (Figure 3.2) show that wild-type and knockout mice are independently clustered, indicating a difference in the expression level dependent on the genotype of the mice. This was expected since the input data for the clustering was already filtered by hypothesis testing. The heat-map suggests that most of the differentially expressed microRNAs are down-regulated in *Ahr*$^{-/-}$ mice compared to *Ahr*$^{+/+}$ mice, while only a few are up-regulated. However, the differences in expression between the groups appear to be rather low. Also, there appears to be very little variation in expression levels of these microRNAs between the sexes. The clearest pattern in the heat-map is seen for miR-592 (located at the bottom of the map) and miR-335-3p (located in the middle) which also here is shown to be opposite regulated.

The results of the PCA showed that wild type and knockout mice were grouped independently. Most microRNAs clustered together, but some microRNAs separated from the others. Of interest is miR-592 and miR-335-3p, which were shown to be op-

posite correlated, in accordance with the fold changes of these two miRNAs.

3.1.4. Selcting miRNAs for further analysis

The four most significantly differentially expressed miRNAs, miR-335-3p, miR-592, miR-33*, and miR-27a*, ($p<0.01$) were chosen for further analysis. All of these miRNAs showed consistent expression patterns in both male and female mice (Figure 3.3, for information about miNRA sequence and chromosome localisation see Appendix D).

3.1.5. Genome database search for AHR response elements

Differences in expression of microRNA's between *Ahr* wild-type and knockout mice could be associated with the occurrence of binding sites for the Ahr transcription complex in regulatory regions upstream of the particular microRNA's. A search for full and core consensus AHR response element I (AHRE-I), and AHRE-II sequences in the promoter region of the four most differentially expressed miRNAs was performed (Mathematica 8, Appendix E). Although few perfect matches with the full AHRE-I and II consensus were found, many core binding regions of AHRE-I were present in both human and mouse. Since no evolutionary analysis was performed, it is not known whether the AHREs identified in the two species are the same ones.

3.1.6. Target prediction and functional pathway analysis

To identify enriched biological processes with mRNA targets of the

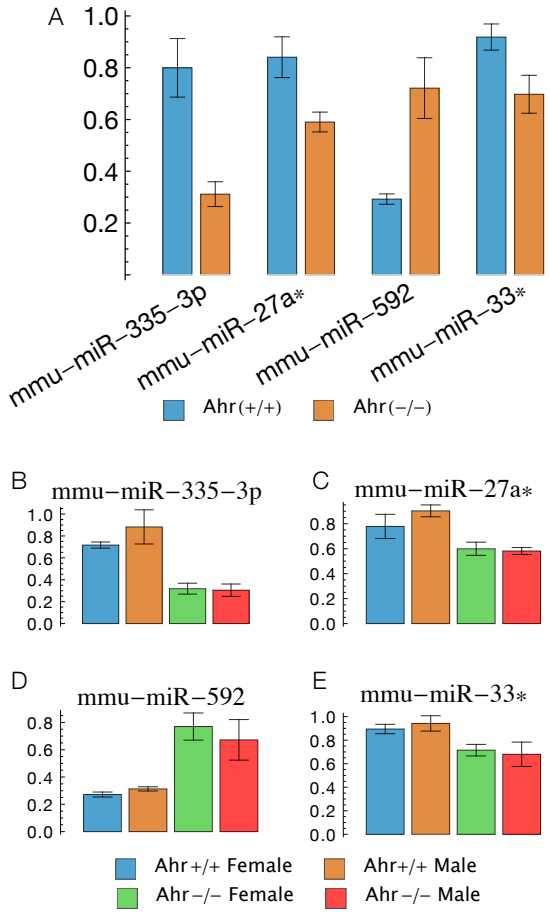


Figure 3.3: Relative expression of the four miRNAs with lowest p-value (*t*-test) in $Ahr^{+/+}$ and $Ahr^{-/-}$ mice. **A:** Average relative expression in male and female combined. Fold changes: miR-335-3p, 0.39 ; miR-27a*, 0.70; miR-592, 2.46; miR-33*, 0.76. **B-E:** Relative expression of each miRNA shown for male and female separately. The graphs show that AHR-dependent regulation is consistent in both sexes.

Table 3.3: A search for AHR response element (AHRE) I and II was doen in Mathematica 8 (Wolfram Research, Inc). Full AHRE-I: 5'-T/GNGCGTGA/CG/CA, Core AHRE-I 5'-GCGTG{Swanson:1995wu}, Full AHRE-II: CATG{N6}CT/ATG{Sogawa:2004il}. mmu: *Mus musculus*, hsa: *Homo sapiens.*The miR-592 was not found by sequence search of the mouse genome in Mathematica. Therefor NA is reported for miR-592 in mouse.

miRNA Name	AHRE-I full (mmu)	AHRE-I core (mmu)	AHRE-II full (mmu)	AHRE-I full (hsa)	AHRE-I core (hsa)	AHRE-II full (hsa)
miR-27a*	2	10	0	0	7	1
miR-33a*	0	8	1	1	13	0
miR-335*	1	10	1	0	7	0
miR-592	NA	NA	NA	0	2	1

Table 3.4: Biological processes where the predicted target mRNA are involved. P-values calculated in MetaCore for both targets of differentially expressed microRNAs and reference microRNAs are shown. A p-value of 1.0 is set for pathways not listed as targets of the reference miRNAs. Specificity ratio is calculated by dividing the p-value of pathways enriched with reference miRNA targets on p-value of pathways enriched with differentially expressed miRNA targets. DE miRNAs: Differentially expressed miRNAs, Ref miRNAs: reference miRNAs selected at random.

Pathway	DE miRNAs	Ref miRNAs	Specificity Ratio
Development_Gastrin in differentiation of the gastric mucosa	3.16E-06	1.0	316456
Signal transduction_cAMP signaling	4.32E-04	1.0	2315
Development_G-Proteins mediated regulation MARK-ERK signaling	6.33E-04	1.0	1580
Development_A2B receptor: action via G-protein alpha s	7.48E-04	1.0	1337
Muscle contraction_ACM regulation of smooth muscle contraction	9.38E-04	1.0	1066
Development_Gastrin in cell growth and proliferation	1.15E-03	1.0	870
Immune response_Gastrin in inflammatory response	1.42E-03	0.163	115

differentially expressed miRNAs, functional ontology enrichment analysis was performed in MetaCore (GeneGo, Thomson Reuters). Ten of the most differentially expressed miRNAs, and ten reference miRNAs chosen at random (by random number generation) from the non-differentially expressed miRNAs were included in the analysis. The predicted mRNA targets of these miRNAs were downloaded from miRWalk, and all mRNAs predicted to be targeted by 3 or more of the chosen microRNAs were uploaded to MetaCore which then provided a p-value for pathways enriched by microRNA predicted targets. A specificity ratio was calculated by dividing the p-value of the pathways found to be coregulated by the reference microRNAs on the p-values of targets coregulated by the differentially expressed microRNAs. Seven pathways were found to be overrepresented (specificity ratio > 100) in the differentially expressed microRNAs.

Most of AHR's known function is in xenobiotic metabolism, although AHR is also involved in several other biological pathways. AHR regulated miRNAs may be important in regulating mRNA targets involved in AHR conducted biological pathways. To investigate the interplay between the four miRNAs chosen for further studies (miR-27a*, miR-33*, miR-335*, and miR-592) and proteins known to be involved in AHR pathways, a pathway map was generated in MetaCore. The four miRNAs, AHR, ARNT, and CYP1A1 were used as input (Figure 3.4).

An interesting notion in this pathway map is that miR-335 directly inhibits the tumour suppressor gene retinoblastoma protein (Rb). It is also worth noticing that the hormonal receptors oestrogen receptor 1 (ESR1) and androgen receptor are involved in this pathway. Especially, ESR1 is indicated as having a inhibitory effect on miR-27a.

3.2.Cell culture experiments

RNA interference experiments, using siRNA targeting *Ahr* mRNA, were performed to investigate the effect of the AHR on seven miRNAs *in vitro*. Four of the miRNAs (hsa-miR-27a*, hsa-miR-33a*, hsa-miR-335*, and hsa-miR-592) were chosen based on the effect of the AHR on expression of their homologues (mmu-miR-27a*, mmu-miR-33*, mmu-miR-335-3p, and mmu-miR-592, respectively) in the knock out mouse experiment. Three additional miRNAs (hsa-miR-122a, hsa-miR-203, and hsa-miR-494) were chosen based on their previously reported dependence of AHR in the literature.

3.2.1.Knock Down of AHR

AHR expression levels were initially studied in five different immortal-

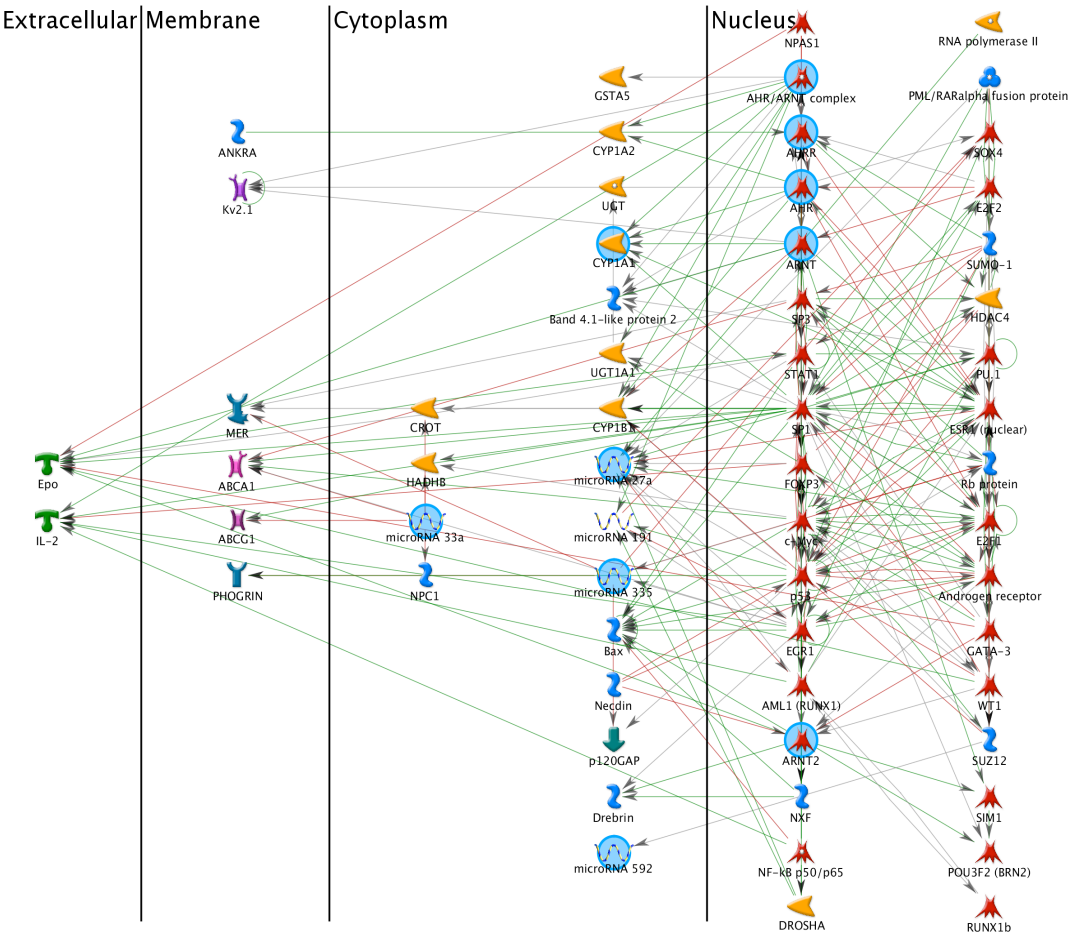


Figure 3.4: A biological pathway map was created in MetaCore using the miR-27a, miR-33a, miR-335, miR-592, AHR, ARNT, and CYP1A1 as input data. The map shows interactions between cellular components. Green arrows mean activation, red lines mean inhibition, and grey lines mean that the function of the interaction is unknown. Explanation of symbols is provided in Appendix H.

ised human lung epithelial cell lines (HBEC2-KT, HBEC3-KT, HBEC12-KT, BEP2D, and BEAS2B). The cells were transfected with three siRNA duplexes, each targeting a unique sequence in the Ahr mRNA without other complimentary sequences in the human genome. The reason for using several siRNA against the same target is to reduce off target effect silencing by a single siRNA. To identify knock down efficiency of each siRNA and that compared to the mixture of all three siRNAs, an experiment was conducted were the siRNAs were transfected separately, and the mRNA levels of AHR was measured by RT-qPCR (figure 3.5). The results showed significant knock down of the AHR. Why the effect of separate siRNAs were meas-

ured in BEP2D cells and not HBEC3-KT is that this was done early in the process, before the cell line for further experiments was chosen. Optimally the this should also have been done in HBEC-3KTs. The stability of the siRNAs and their knock down effect on AHR was studied in HBEC3-KT over 0, 1, 3, and 5 days after replacing transfection medium with LHC-9 medium (for 0 days the cell line experiment was stopped immediately after removing transfection medium). The results showed no reduction of silencing potential over time.

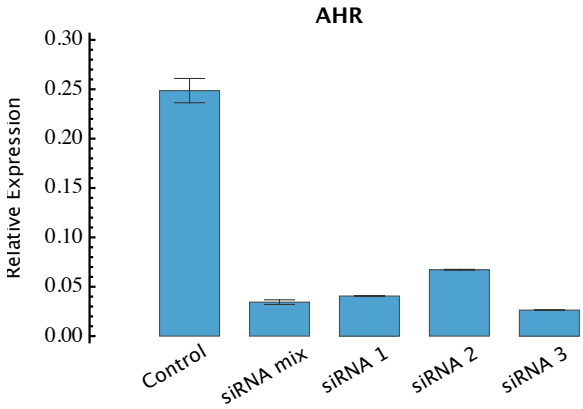


Figure 3.5: BEP2D cells were transfected with control siRNA, a mixture of three siRNAs targeting AHR mRNA, and each of the three siRNAs independently. In each case n=3. Expression levels of *AHR* was measured by RT-qPCR. Relative expression levels were calculated using the PCR-efficiency adjusted ΔC_q method. Fold change (relative to control) for siRNA mix, siRNA 1, siRNA 2, and siRNA 3, respectively, are as follows: **AHR**: 0.14. 0.16. 0.27. 0.11.

3.2.2. Cell morphology after siRNA transfection

The cells transfected with AHR siRNA were observed to have a different phenotype than cells transfected with control siRNA in the phase contrast light microscope (Figure 3.6). The AHR siRNA transfected cells appeared bigger and more flattened than their control transfected and non-transfected counterparts. Also, AHR siRNA transfected cells contained more vacuoles and more cell-cell conjunctions.

3.2.3. Toxicity of AHR siRNA transfection

To test for toxicity of the AHR siRNAs and the control siRNAs a toxicity assay using cell-titer blue was carried out in two different cell lines (HBEC3-KT, BEP2D). No toxic effect of neither AHR siRNA nor control siRNA was detected in any of the two cell lines (Figure 3.7).

3.2.4. MicroRNA expression

To examine the effect of AHR on miRNAs *in vitro* a knock down experiment was performed on HBEC3-KT. Cells were transfected with either AHR siRNA or control siRNA and subsequently exposed for B[a]P, CSC, or DMSO for 1 or 5 days. microRNA expression

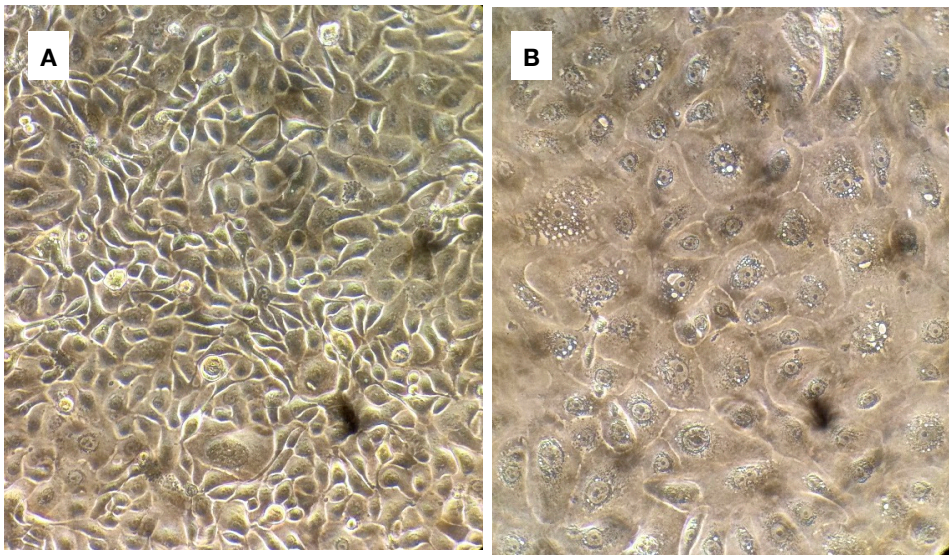


Figure 3.6: A: HBEC3-KT transfected with control siRNA. **B:** HBEC3-KT transfected with AHR siRNA.

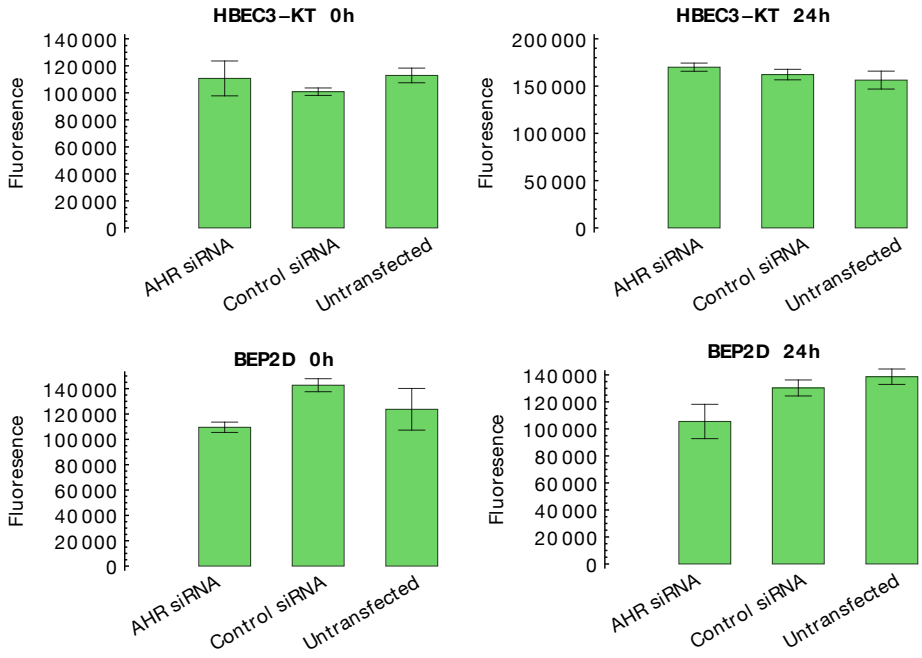


Figure 3.7: Toxicity test using cell titer blue revieled no significant toxicity of neither control siRNA nor AHR siRNA. 0 h: Cell titer blue assay immediatly after transfection medium was removed. 24 h: after removing transfection medium, cells were incubated for another 24 hours before cell titer blue assay was carried out. Background fluoresence is subtracted from the shown values.

levels was measured at two time points, one and five days of exposure. The experiment was repeated three times independently, with two parallel's of each sample.

AHR and CYP1A1 expression were measured for each sample by RT-qPCR to control for knockdown effect of siRNA. CYP1A1 mRNA level was used as a control for reduced capacity of AHR mediated induction (Figure 3.8). In addition, ARNT expression was measured in one of the experiments.

For the miRNAs, total RNA from the two parallel's within each experiment were combined before cDNA synthesis, giving three biological replicates for each exposure at each time point.

The difference in AHR mRNA levels between AHR siRNA transfected and control cells varied from each experiments, with fold changes ranging from as little as 1.26 to 22.19. CYP1A1 was found to be strongly induced by exposure of B[a]P or CSC as compared to control cells, AHR resulted in obliteration of CYP1A1 expression in all cells. The ARNT was not significantly affected by AHR silencing in this experiment.

Expression levels of seven miRNAs were measured, three selected from previous reports, and four selected from the AHR knock out mice experiment in this theses. No significant differences were found in the four miRNAs chosen from the knock out experiment (Figure 3.9). For the miRNAs previously described in the literature, miR-122a and miR-203 were significantly down-regulated in AHR siRNA cells compared to control (Figure 3.10). miR-203 was also significantly down-regulated in cells exposed to B[*a*]P or CSC for five days compared to control (DMSO).

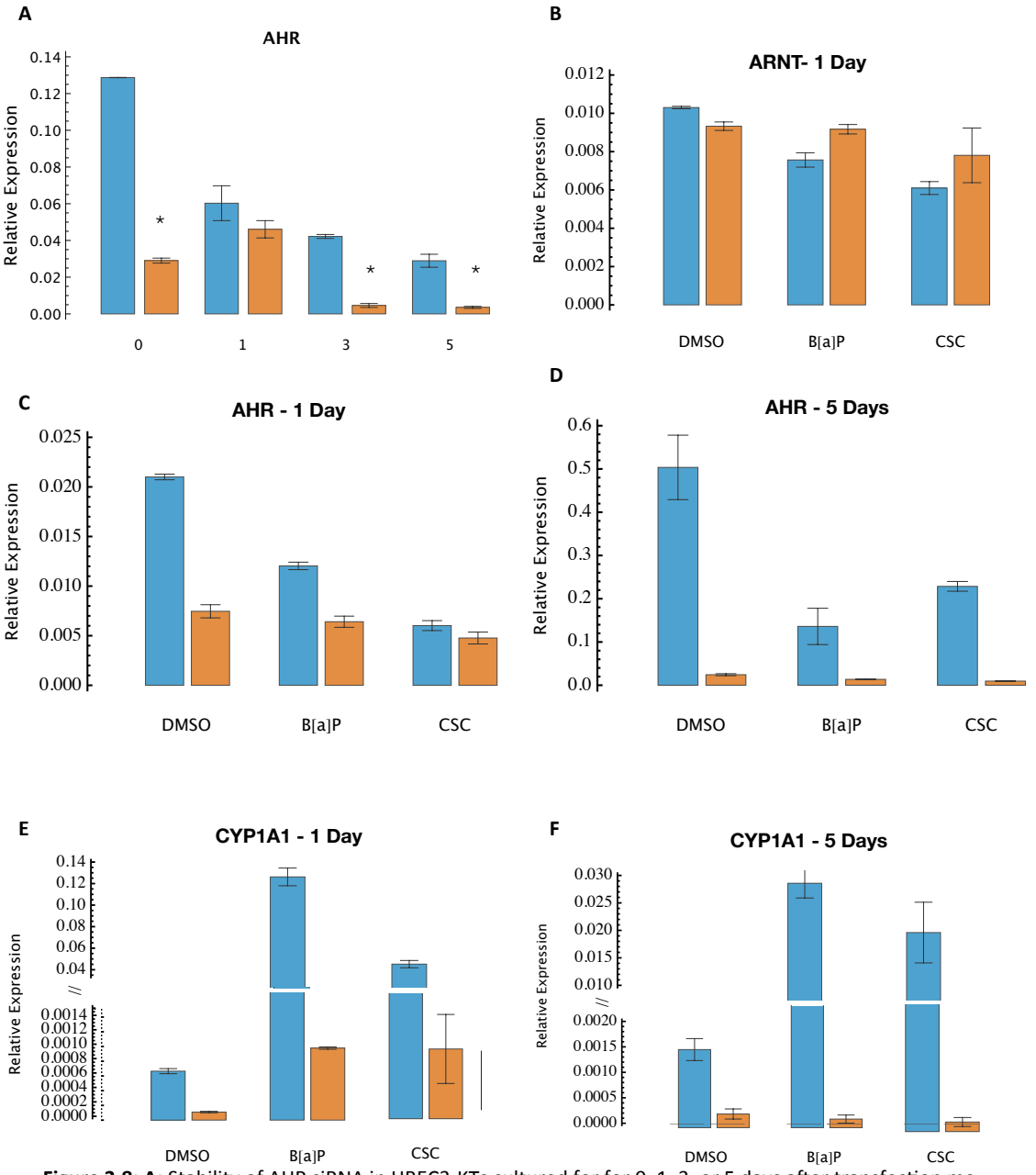


Figure 3.8: **A:** Stability of AHR siRNA in HBEC3-KTs cultured for for 0, 1, 3, or 5 days after transfection medium is replaced with LHC-9 medium. Fold changes are: 0 days: 0.23 (p=6.14x10-6), 1 day: 0.76 (p=0.21), 3 days: 0.11 (p=4.7x105), 5 days: 0.12 (p=0.0024). Significant knockdown (t-test, p<0.05) is indicated by an asterisk. **B-F**: Expression levels of *ARNT, AHR*, and *CYP1A1* in HBEC3-KTs transfected with AHR siRNA or control siRNA. Note that the lower and upper half of the CYP1A1 axes have different dimensions. The fold changes for DMSO, B]a]P, and CSC, respectively, is: **ARNT:** 0.9 (p=0.051), 1.2 (p=0.07), 1.3 (p=0.36) **AHR:** 2.8 (p=0.003), 1.9 (p=014), 1.3 (p=0.25), **CYP1A1:** 6.1 (p=0.004), 126.6 (p=0.043) 44.4 (p=0.047).

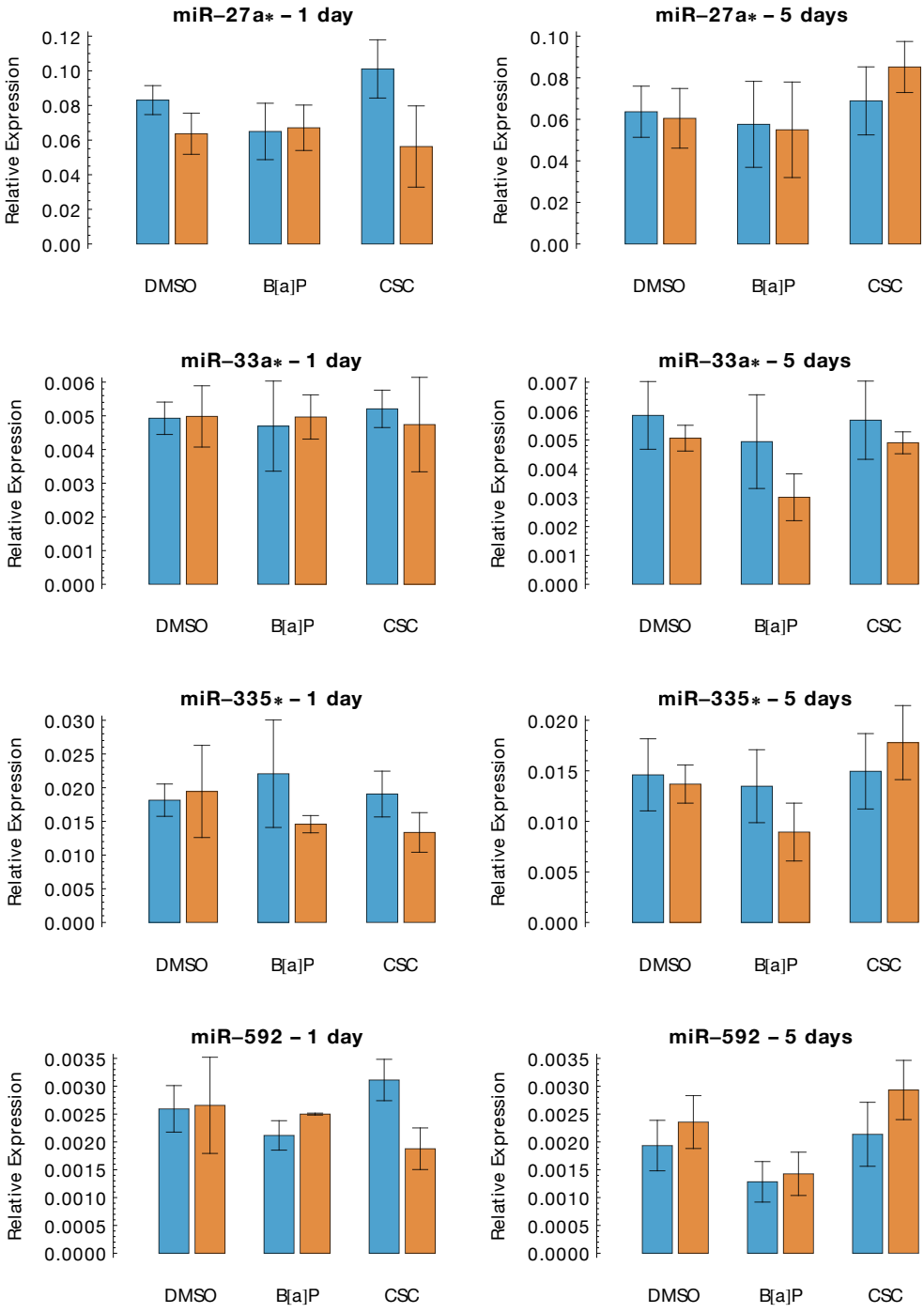


Figure 3.9: Four differentially expressed miRNAs were chosen from the knock out mice experiment and their expression levels were measured in HBEC3-KTs by RT-qPCR. None of the miRNAs were found to be significantly ($p<0.5$, 3-way ANOVA) affected by neither AHR silencing, exposure, duration, nor any combination of the treatments.

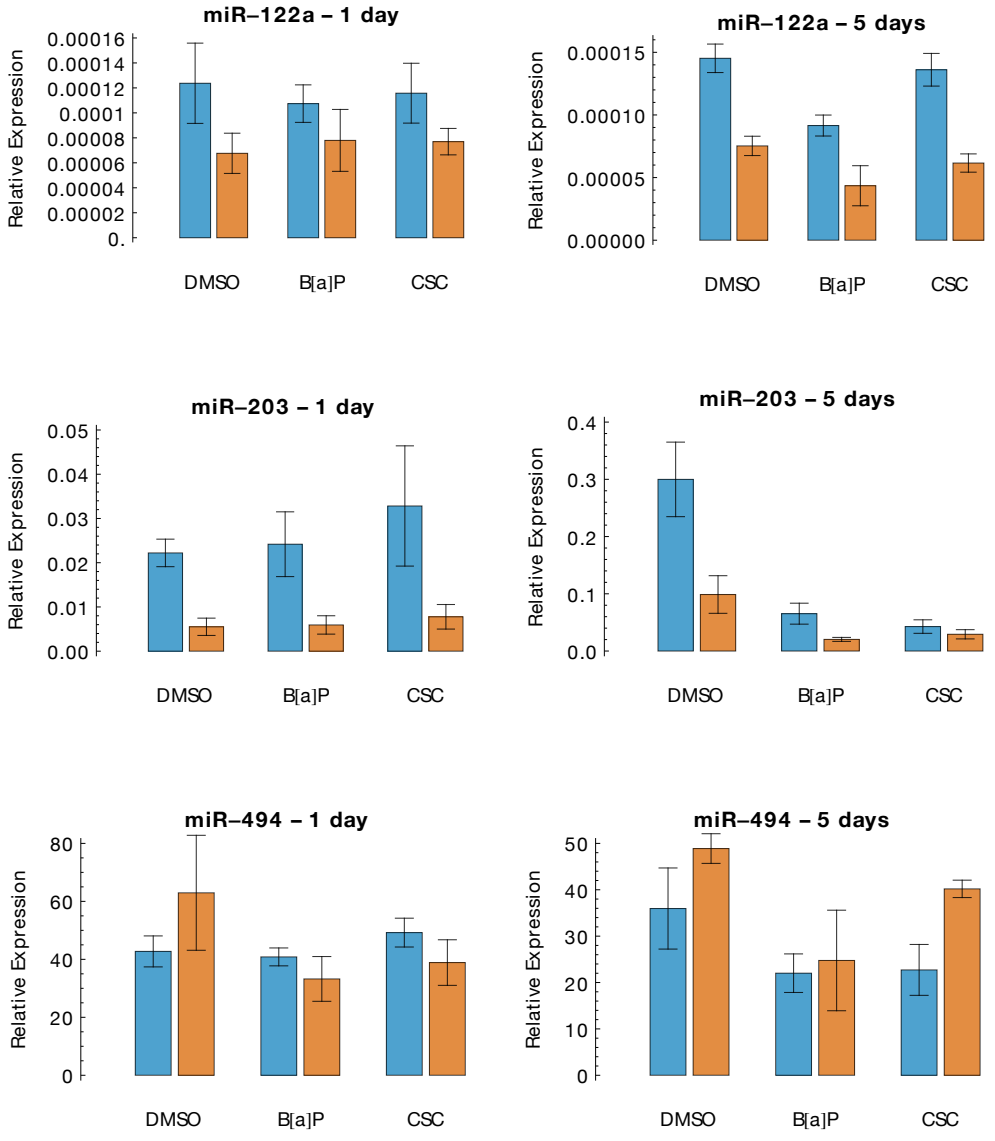


Figure 3.10: The expression levels of three miRNAs previously found to be affected by exposure of B[*a*]P were investigated in HBEC3-KTs. Expression of miR-122a was found to be significantly down regulated (p=0.00012) in AHR siRNA transfected cells. miR-203 was found to be significantly affected by siRNA transfection (control vs. AHR siRNA, p=0.0015), exposure (B[a]P and CSC compared to DMSO, p=0.0002), and by duration of exposure (p=0.00003). miR-494 expression was significantly altered by exposure (B[a]P compared to DMSO, p=0.023), and by duration of exposure (p=0.018).

4.Discussion

Little is known about epigenetic mechanisms in toxic effects of environmental pollutants. MicroRNAs are involved in many, if not all, biological processes and function as down-regulators of gene expression by sequence specific targeting of mRNAs. Aberrant expression of several miRNAs has been observed in different types of cancer, including lung cancer[113,88]. Emerging evidence suggest that miRNA expression can be regulated by environmental factors[84]. How environmental factors influence epigenetic changes is still largely unclear. Increased knowledge of the mechanism by which environmental compounds lead to changes in the epigenome will lead to increased insight into their possible carcinogenic effects.

In this thesis, the effect of the AHR on miRNA expression was examined. The most studied and most understood biological pathway in which AHR is involved, is xenobiotic metabolism. AHR binds exogenous planar polyaromatic ligands, such as PAHs, which among others are present in tobacco smoke, and dioxin like compounds. Ligand binding to the AHR leads to an activated transcription factor that ultimately induces transcription of among others phase I and phase II xenobiotic metabolising genes. Increasing evidence suggests that the AHR is involved in a wide range of biological processes including development, cell cycle regulation, cell signalling, cytoskeleton organisation, extra cellular matrix (ECM) modification, adhesion, immune responses, and circadian rhythm[34,40-43]. Accumulating evidence for importance of the AHR in a diverse range of biological processes underlines the need to identifying miRNAs regulated by this transcription factor. Since miRNAs are transcribed by RNA pol II, they are classified as class II genes, and can be transcriptionally regulated by transcription factors that regulate transcription of this class of genes, such as the AHR. We hypothesised that the AHR may mediate some of its biological function by regulating the expression of certain miRNAs, and aimed at identifying such miRNAs.

An initial high-throughput RT-qPCR screening of the expression of 750 miRNAs in mouse lung was performed to identify candidate miRNAs whose expression potentially is regulated by AHR. Fifteen miRNAs were

found to show significantly different expression in *Ahr*$^{-/-}$ compared to *Ahr*$^{+/+}$ mice. Although the fold change for most of the miRNAs was relatively low, the biological consequences could be significant due to each miRNA being able to target several mRNAs[86]. In addition, if the miRNAs regulated by AHR act on the same biological pathways the cumulative effect could be of significance.

To identify pathways that may be co-regulated by differentially expressed miRNAs in the knockout mouse experiment, a functional ontology enrichment analysis was performed in MetaCore (GeneGo, Thomson-Reuters). Seven biological processes were found to be enriched with targets of 10 selected miRNAs. Interestingly, six of these pathways are related to G-protein mediated signalling. As mentioned previously (section 1.3.2), G-protein mediated signalling can negatively regulate AHR activity. G-protein coupled receptors (GPCR) play an important role in cancer progression and metastasis. Disruption of normal GPCR functionality in tumours can lead to autonomous cell proliferation, evading immune cell detection, increase in nutrient and oxygen supply, invading of surrounding tissues, and dissemination to other organs[114].

Gastrin releasing peptide (GRP) stimulates cell proliferation and growth through the GPCR family member GRP receptor (GRPR). GRP has been shown to be important in tumourigenesis and is overexpressed in lung tumors[25,115]. An increase in expression of GRPR mRNA in human lung has been shown to be associated with smoking, and even more so in female smokers than in male smokers[116].

Since there seem to be a competing role between TCDD and G-protein activated AHR[38,39], it may be hypothesized that miRNAs can mediate repression of mRNAs involved in G-protein signalling in the presence of exogenous ligands, in order to suppress G-protein mediated inhibition of AHR. This would make more AHR available for ligand binding and XME induction, increasing metabolic capacity.

All but one of the differentially expressed miRNAs were down regulated. The AHR/ARNT complex induces transcription by binding to AHR responsive elements (AHRE I or AHRE II). The down regulation of certain miRNAs in the absence of AHR, suggests that these miRNAs might be transcriptionally induced by direct binding of AHR/ARNT complex to AHRE in their promoter region. A loss of AHR functionality would then lead to reduced expression of these miRNAs. Another explanation would be that transcription of these miRNAs is induced by other transcription factors

downstream of AHR in the AHR pathway.

Four miRNAs from the mouse experiment (miR-27a*, miR-33*, miR-335-3p, and miR-592), were chosen for further studies in human lung cell lines. The three first mentioned miRNAs were found to be down-regulated in *Ahr*$^{-/-}$ mice compared to *Ahr*$^{+/+}$ mice. To determine if the AHR-dependent regulation could be explained by AHRE located in the promoter region of these three miRNAs, a search for full and core AHRE-I and full AHRE-II sequences was performed. Few full AHRE-I or AHRE-II were found in the human and mouse genome in promoter regions of these miRNAs. However, many core AHRE-II sequences were found and some of these differed in as little as one nucleotide from the full sequence. If the AHR/ARNT-complex is able to bind to the core sequence without full consensus similarity of the flanking sites, there is a possibility that one or more of the elements found could be responsible for inducing transcription of one of the microRNAs. For these miRNAs, however, this has not been demonstrated experimentally.

The AHRE-I core sequence is five nucleotides long which means that if randomly distributed in the genome, it would be expected to occur once every kilo base. A search for AHRE-I core sequence in the entire chromosome 6 (126 Mb) in the mouse genome, for example, results in only 28,000 AHRE-I core sequences found, giving an expected occurrence every fourth kilo base. This suggests that since the five nucleotide AHRE-I core sequence has a biological function, it has been selectively conserved in regions close to genes that depend on AHR for appropriate transcription, while at the other hand, selectively removed through evolution from genomic regions where it may result in aberrant transcription of nearby genes.

The only miRNA upregulated in *Ahr*$^{-/-}$ mice compared to *Ahr*$^{+/+}$ mice was miR-592. Since AHR functions as a transcriptional inducer by direct binding of response elements upstream of transcription start sites of its target genes, it is unlikely that transcription of this miRNA is directly regulated by AHR. The induction of mir-592 transcription may be caused transcriptional up-regulation by other bHLH/PAS family proteins that might compete with AHR for dimerisation to ARNT. It may be hypothesised that in the absence of AHR, ARNT will then be free to dimerise with these transcription factors, thus leading to an increased expression of miR-592.

miR-27a* was found to be significantly down-regulated in both male and female *Ahr*$^{-/-}$ mice. This miRNA has previously been reported to be significantly up-regulated in SCLC[117],

gastric cancer[118], and prostate cancer[119], and down-regulated in multidrug resistant human ovarian cancer cells[120]. The indication of miR-27a being deregulated in several types of cancer suggests that it might be of importance for tumour cells to either obtain or sustain their cancerous state.

miR-27a* is found to target the mRNA of a zinc finger protein, ZBTB10, that acts as a repressor of specificity proteins (Sp)[121]. The Sp-1 transcription factor is involved in inducing transcription of oestrogen receptor 1 (ESR1, also known as ER-α), thus miR-27a* mediated repression of the Sp-1 repressor results in induced expression of ESR1 protein[122]. The miR-27a has been found to be down-regulated by estradiol (E2), caused by direct inhibition of *mir-27a* transcription by ESR1. Four ESR1 binding sites are found in the promoter region of the *mir-23a~mir24~mir27a* gene cluster[123]. Oestrogen may be involved in lung carcinogenesis. Oestrogen receptors have been associated with an increased metabolic activation of PAHs in lung tissues, and there is evidence for cross-talk between ESR and AHR [124,125]. Lung tumors from women are more likely to express ESR than lung tumours from men, and hence may account for some of the sex differences in susceptibility to lung cancer[25]. The cross-talk between AHR and ESR pathways may involve miRNAs, and miR-27a* may be a putative mediator of this cross-talk.

Little literature exists on neither mmu-miR-335-3p, nor its human homologue hsa-miR-335*. In contrast, a few studies have been performed for the hsa-miR-335 (major) sequence. Expression of mir-335 has been found be significantly correlated with transcription of its host gene MEST, indicating a dependent transcriptional regulation of these two genes[126]. The possible Ahr dependent regulation of miR-335-3p found in this thesis is interesting in relation to previous reports where it was up-regulated in a TP53 dependent manner as a response to DNA damage[127]. The RB1 mRNA is directly targeted and down-regulated by miR-335, resulting in cell cycle arrest[127]. This hints at a possibility that in the presence of carcinogenic AHR ligands, the AHR may induce miR-335 transcription in order to rapidly down-regulate RB1 mRNA which gives DNA repair enzymes time to remove DNA adducts before cell cycle progression and DNA replication. The AHR may regulate expression of miR-335 directly by AHRE binding, or indirectly through crosstalk with TP53. However, how these hypothesis might relate to miR-335-3p/miR-335* is not known.

The *mir-33a* is located in an intron of the sterol regulatory element binding transcription factor 2 (*SREBF2*), a transcription factor that controls cho-

lesterol biosynthesis and uptake by stimulating transcription of sterol-regulated genes. Expression of *mir-33* is co-regulated with the expression of *SREBF2*, and the miRNA functions in concert with its host gene to regulate cholesterol homeostasis[128]. Down-regulation of miR-33 has been observed in lung cancer[129]. The AHR-dependent regulation observed in the mouse experiment in this thesis suggests that miR-33* may mediate AHR-dependent tumour suppression.

No studies exist on the specific function of miR-592. Two of the highest scoring predicted targets of this miRNA are involved in immune system functions. Glia maturation factor β (GMFB) is involved in T-cell maturation[130], while tumour necrosis factor (ligand) superfamily member 18 (TNFSF18) is involved in T-lymphocyte survival in peripheral tissues. How this might relate to lung cancer is not known.

The PCA and clustering analysis grouped the wildtype and knockout mice separately. However, the dataset used for these analyses was already filtered by statistical testing (*t*-test and ANOVA), so the observed grouping is not evidence of global effects, but rather expected from the input dataset. Even so, the PCA gives information of outlier miRNAs different from the major group and their correlation. The heatmap emphasise the downregulation of the majority of differentially expressed miRNAs.

The expression levels of the four miRNAs chosen for further studies were investigated in an RNA interference experiment. Human bronchial epithelial cells (HBEC3-KT), a lung cell line immortalised by over-expression of *CDK4* and h*TERT* showing a "normal" phenotype[100], was transfected with siRNAs specific for AHR receptor[100]. Cell lines may not be an ideal model system due to disadvantages like genetic instability, selective growth of subpopulations during culturing, and lack of microenvironment and other *in vivo* interactions. However, cell lines are invaluable as a complement to animal studies, and can provide for instance provide invaluable knowledge of mechanisms of cancer origins[131].

The HBEC3 cell line has been reported to be cultured for more than 240 populations without losing normal phenotype, demonstrating its stability over several division cycles[131]. This is in contrast to cancer cell lines where it can be expected to find thousands of mutations and a high degree of genome instability that might lead to rapidly changing phenotypes[132]. One concern by using HBEC3-KT in the work for this thesis is that CDK4 is known to form a complex with AHR, and might be competing with ARNT for AHR binding. This would result in reduced

expression of AHR/ARNT complex target genes.

Transfection of AHR specific siRNA resulted in stable knock down of AHR mRNA for at least 5 days. Although the AHR protein level was not measured, *CYP1A1* expression and induction was almost completely ablated in AHR siRNA transfected cells, suggesting a significantly reduced level of AHR transcriptional inducing activity in these cells. The knockdown was not complete, resulting in a low level of induction of *CYP1A1* also in transfected cell, however, this was to a significantly lesser extent than in untransfected cells. The mRNA level of AHR decreased in B[*a*]P and CSC exposed cells. This agree with previous studies were reduction of AHR protein have been observed upon B[*a*]P[133].

Along with the four miRNAs from the mouse experiment, three additional miRNAs, miR-122a, miR-203, and miR-494, were selected from the literature based on reported AHR-dependent transcriptional induction *in vitro* (see section 2.2). In the RNA interference experiment in this thesis, miR-122a and miR-203 were shown to be down-regulated in AHR knockdown cells. The expression of miR-494 was not significantly altered by AHR siRNA transfection, however, the Cq values of this miRNA were very low (<15), meaning that a lower amount of RNA should optimally have been used for the RT-qPCR. Unfortunately, lack of time did not allow for a rerun of this miRNA for this thesis.

Although miR-122a and miR-203 were downregulated in the absence of AHR, exposure to B[*a*]P and CSC did not significantly induce transcription of these miRNAs. As mentioned in section 1.3.2, G-proteins can directly sequesters AIP from the AHR complex, which result in ubiquitination and nuclear translocation of AHR[39]. Although this was found to result in AHR having an inhibitory function on *CYP1A1* induction, it is unknown whether the ubiquitinated AHR can form complex with other bHLH/PAS family members and induce transcription from other response elements than those bound by AHR/ARNT complex. Different ligands of AHR have been shown to induce transcription of different target genes[134]. This observation may be caused by AHR dimerization to different bHLH/PAS dependent on the bound ligand. AHR's ability to bind different partners in the transcription regulatory complex may be due to conformational changes in AHR protein structure caused by ligands with slightly different structure [135]. It might be speculated that induction of miR-122a and miR-203 may be dependent on other exogenous or endogenous ligands, and AHR dimerising with other proteins than ARNT.

Levels of miR-203 was significantly lower in B[*a*]P and CSC exposed cells after five days compared to control cells. This can be explained by the down-regulation of AHR mRNA observed in exposed cells. If transcription of this miRNA is induced in an ARNT independent manner, it is also possible that introduction of xenobiotics result in competitive dimerisation with ARNT.

The four differentially expressed miRNAs in *Ahr*$^{+/+}$ and *Ahr*$^{-/-}$ mice chosen for further validation experiments, were not found to be differentially expressed in HBECs *in vitro*. The reason for this lack of correspondence could be due to differences between mouse and human experimental systems, however, the miRNAs chosen from the literature were also from mouse data and these showed similar results in human cells. The previous work on the three miRNAs from the literature was *in vitro* data. This might suggest that the reason for lack of reproducibility between mice and human in this thesis could be due to differences between *in vivo* and *in vitro* situations, rather than species differences.

The observed differences in miRNA expression between *Ahr* knockout and *Ahr* wild type mice are identical in male and female individuals, strongly supporting that these findings are real. Moreover, the biological processes that the identified miRNAs are involved in can be tied to AHR function, corroborating the notion that these miRNAs may be involved in mediating some of the biological effects of the Ahr *in vivo*.

The cells transfected with AHR siRNA were observed to have a different phenotype than cells transfected with control siRNA in the phase contrast light microscope. The AHR siRNA transfected cells appeared bigger and more flattened than their control transfected and non-transfected counterparts. Also, AHR siRNA transfected cells contained more vacuoles and more cell-cell conjunctions. This is in accordance with previous findings that *Ahr*$^{-/-}$ murine fibroblasts exhibit a more spread morphology, associated with prominent actin stress fibres and to increased numbers of focal adhesions with a depolarised distribution around the cell area[40]. The AHR is involved in altering the extracellular matrix (ECM) and adhesion, and is associated with enhance migration and invasiveness[42].

The small (or lack of) effect of exposure on the expression of miRNAs *in vitro* is consistent with previous reports on the effect of B[a]P and TCDD on miRNA expression. This could suggest that epigenetics, and miRNAs specifically, are not important in mediating toxic effects of PAHs. However, diesel exhaust particles (DEP), of which the organic adsorbents (and particular PAHs) are considered to play the important role in toxicity, have been re-

ported to induce considerable changes on miRNA expression. Jardim et al. found that 130 miRNAs, including miR-494, had increased expression higher than 1.5-fold, and 67 miRNAs, including miR-27a, had more than a 1.5-fold decrease in expression upon exposure to DEP in human lung cells *in vitro*[136]. DEPs adsorb different types of toxic compounds, including PAHs. It is possible that regulation of miRNA by DEP (or PAH on the particles) could both be a result of AHR activation, but possibly by different ligands forming different transcription factor complexes [134].

Further studies

The mouse experiment was just performed once so it would be interesting to rerun the TaqMan array on another set of *Ahr*$^{+/+}$ and *Ahr*$^{-/-}$ mice to verify the reproducibility of the data. Also, it would be interesting to profile miRNA expression in lung from mice exposed to B[*a*]P and/or CSC. This will provide information whether the same miRNAs found to be AHR dependent in this thesis are also induced by exposure, or if exogenous ligand activated AHR induce transcription of a different set of miRNAs. This information can provide important insight into how AHR may mediate some of the carcinogenic effect of xenobiotics, either by induction of a new set of miRNAs, or by disruption of intrinsic endogenous AHR function.

In this thesis miRNA expression was only studied in one human lung epithelial cell line. Further studies should include one or more additional lung cell lines, for example immortalised by different means, and could also include tumour cell lines. Since AHR dependent miRNA expression may differ *in vitro* and *in vivo*, it would be interesting to run expression profiling arrays on cell lines as well. The time span of exposure could be extended to see if miRNA expression changes over periods longer than five days. Different dosages of PAH/CSC would also be of interest. Additionally, it would be very interesting to expose the cells to other PAHs than B[*a*]P, for instance nitro-PAHs, to see if different AHR ligands can induce expression of different sets of miRNAs.

The next step would be to assay the functionality of the identified miRNAs. This could be done either by screening of mRNA expression, or by RT-qPCR of some of the predicted targets of each miRNA. It would be natural to include miRNA inhibitors and miRNA mimics in these assays, knocking down or over-expressing miRNA activity, respectively.

Whether the AHRE identified in this thesis are functional is not known. Chip-on-Chip studies may be performed to find out whether any of the AHRE elements in the promoter region

of the miRNAs are true AHR/ARNT complex binding elements.

5.Conclusion

The aim of this thesis was to identify miRNAs potentially regulated by AHR. This was done by initial miRNA expression profiling in lung of *Ahr* wild type and *Ahr* knockout mice. The biological pathways in which the identified miRNAs may be involved were studied by a bioinformatic approach. Four miRNAs from the mouse experiment, whose expression were most significantly divergent, together with three miRNAs previously reported to be regulated by the *Ahr* in murine cells, were studied in an RNA interference (RNAi) experiment in immortalised human lung epithelial cell lines. It was shown that expression of several miRNAs possibly is regulated in an AHR-dependent manner. Discrepancies between *in vivo* and *in vitro* experiments suggest that care must be taken when extrapolating miRNA expression data from cell culture studies to whole organs or organisms.

References

1. Jemal, A. et al. Global cancer statistics. *CA: A Cancer Journal for Clinicians* **61**, 69-90 (2011).

2. *GLOBACON 2008*. *GLOBACON 2008* at <http://globocan.iarc.fr/>

3. Youlden, D.R., Cramb, S.M. & Baade, P.D. The International Epidemiology of Lung Cancer: geographical distribution and secular trends. *J Thorac Oncol* **3**, 819-831 (2008).

4. IARC *Tobacco smoke and involuntary smoking.* **83**, (IARC Press: 2004).

5. Hecht, S.S. Cigarette smoking: cancer risks, carcinogens, and mechanisms. *Langenbecks Arch Surg* **391**, 603-613 (2006).

6. Yenugadhati, N., Birkett, N., Momoli, F. & Krewski, D. Occupations and Lung Cancer: A Population-Based Case-Control Study in British Columbia. *J. of Toxicology & Env. Hlth., Part A* **72**, 658-675 (2009).

7. Matsumoto, Y. et al. Aryl hydrocarbon receptor plays a significant role in mediating airborne particulate-induced carcinogenesis in mice. *Environ. Sci. Technol.* **41**, 3775-3780 (2007).

8. Vineis, P., Forastiere, F., Hoek, G. & Lipsett, M. Outdoor air pollution and lung cancer: recent epidemiologic evidence. *Int. J. Cancer* **111**, 647-652 (2004).

9. Cohen, A. Outdoor air pollution and lung cancer. *Environ. Health Perspect.* **108**, 743-750 (2000).

10. Schwartz, A.G., Prysak, G.M., Bock, C.H. & Cote, M.L. The molecular epidemiology of lung cancer. *Carcinogenesis* **28**, 507-518 (2007).

11. Matakidou, A., Eisen, T. & Houlston, R.S. Systematic review of the relationship between family history and lung cancer risk. *Br. J. Cancer* **93**, 825-833 (2005).

12. Subramanian, J. & Govindan, R. Lung cancer in never smokers: A review. *J Clin Oncol* **25**, 561-570 (2007).

13. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).

14. Hanahan, D. & Weinberg, R.A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646-674 (2011).

15. Brennan, P., Hainaut, P. & Boffetta, P. Genetics of lung-cancer susceptibility. *Lancet Oncol.* **12**, 399-408 (2011).

16. Bartsch, H., Haugen, A., Risch, A., Shields, P. & Vineis, P. Tumours of the lung: Genetic susceptibility. *Pathology and Genetics of Tumours of the Lung, Pleura,Thymus and Heart* (2004).at <http://www.iarc.fr/en/publications/pdfs-online/pat-gen/bb10/index.php>

17. Kang, J.U., Koo, S.H., Kwon, K.C., Park, J.W. & Kim, J.M. Gain at chromosomal region 5p15.33, containing TERT, is the most frequent genetic event in early stages of non-small cell lung cancer. *Cancer Genet. Cytogenet.* **182**, 1-11 (2008).

18. Chopra, M. & Schrenk, D. Dioxin toxicity, aryl hydrocarbon receptor signaling, and apoptosis-Persistent pollutants affect programmed cell death. *Crit. Rev. Toxicol.* **41**, 292-320 (2011).

19. Zienolddiny, S. et al. Polymorphisms of the interleukin-1 beta gene are associated with increased risk of non-small cell lung cancer. *Int. J. Cancer* **109**, 353-356 (2004).

20. Campa, D. Association of a common polymorphism in the cyclooxygenase 2 gene with risk of non-small cell lung cancer. *Carcinogenesis* **25**, 229-235 (2003).

21. Xu, C., Li, C. & Kong, A. Induction of phase I, II and III drug metabolism/transport by xenobiotics. *Arch Pharm Res* **28**, 249-268 (2005).

22. Taioli, E. Gene-environment interaction in tobacco-related cancers. *Carcinogenesis* **29**, 1467-1474 (2008).

23. Timofeeva, M.N. et al. CYP450 polymorphisms as risk factors for early-onset lung cancer: gender-specific differences. *Carcinogenesis* **30**, 1161-1169 (2009).

24. D, M.C.M. et al. Single Nucleotide Polymorphisms in Tobacco Metabolism and DNA Repair Genes and Prognosis in Resected Non-Small-Cell Lung Cancer. *Journal of Surgical Research* **167**, e5-e12 (2011).

25. Kiyohara, C. & Ohno, Y. Sex differences in lung cancer susceptibility: a review. *Gend Med* **7**, 381-401 (2010).

26. Uppstad, H. et al. Sex differences in susceptibility to PAHs is an intrinsic property of human lung adenocarcinoma cells. *Lung Cancer* **71**, 264-270 (2011).

27. Mollerup, S. et al. Sex differences in risk of lung cancer: Expression of genes in the PAH bioactivation pathway in relation to smoking and bulky DNA adducts. *International Journal of Cancer* **119**, 741-744 (2006).

28. Haugen, A. Women who smoke: are women more susceptible to tobacco-induced lung cancer? *Carcinogenesis* **23**, 227-229 (2002).

29. Bollati, V. & Baccarelli, A. Environmental epigenetics. *Heredity* **105**, 105-112 (2010).

30. Savage, S.A. & Chanock, S.J. Genetic association studies in cancer: good, bad or no longer ugly? *Hum. Genomics* **2**, 415-421 (2006).

31. Poland, A., Glover, E. & Kende, A. Stereospecific, high affinity binding of 2,3,7,8-tetrachlorodibenzo-p-dioxin by hepatic cytosol. Evidence that the binding species is receptor for induction of aryl hydrocarbon hydroxylase. *Journal of Biological Chemistry* **251**, 4936-4946 (1976).

32. Denison, M.S., Pandini, A., Nagy, S.R., Baldwin, E.P. & Bonati, L. Ligand binding and activation of the Ah receptor. *Chem. Biol. Interact.* **141**, 3-24 (2002).

33. Wei, Y.D., RANNUG, U. & Rannug, A. UV-induced CYP1A1 gene expression in human cells is mediated by tryptophan. *Chem. Biol. Interact.* **118**, 127-140 (1999).

34. Furness, S.G.B., Lees, M.J. & Whitelaw, M.L. The dioxin (aryl hydrocarbon) receptor as a model for adaptive responses of bHLH/PAS transcription factors. *FEBS Lett.* **581**, 3616-3625 (2007).

35. Furness, S.G.B. & Whelan, F. The pleiotropy of dioxin toxicity - Xenobiotic misappropriation of the aryl hydrocarbon receptor's alternative physiological roles. *Pharmacology & Therapeutics* **124**, 336-353 (2009).

36. Swanson, H.I., Chan, W.K. & Bradfield, C.A. DNA binding specificities and pairing rules of the Ah receptor, ARNT, and SIM proteins. *Journal of Biological Chemistry* **270**, 26292-26302 (1995).

37. Sogawa, K. et al. A novel induction mechanism of the rat CYP1A2 gene mediated by Ah receptor-Arnt heterodimer. *Biochem. Biophys. Res. Commun.* **318**, 746-755 (2004).

38. Oesch-Bartlomowicz, B. et al. Aryl hydrocarbon receptor activation by cAMP vs. dioxin: divergent signaling pathways. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 9218-9223 (2005).

39. Nakata, A. et al. G-protein signalling negatively regulates the stability of aryl hydrocarbon receptor. *EMBO Rep.* **10**, 622-628 (2009).

40. Carvajal-Gonzalez, J.M. et al. The dioxin receptor regulates the constitutive expression of the vav3 proto-oncogene and modulates cell shape and adhesion. *Mol. Biol. Cell* **20**, 1715-1727 (2009).

41. Dietrich, C. & Kaina, B. The aryl hydrocarbon receptor (AhR) in the regulation of cell-cell contact and tumor growth. *Carcinogenesis* **31**, 1319-1328 (2010).

42. Kung, T., Murphy, K. & White, L. The aryl hydrocarbon receptor (AhR) pathway as a regulatory pathway for cell adhesion and matrix metabolism. *Biochemical Pharmacology* **77**, 536-546 (2009).

43. Shimba, S. & Watabe, Y. Crosstalk between the AHR signaling pathway and circadian rhythm. *Biochemical Pharmacology* **77**, 560-565 (2009).

44. Nguyen, L.P. & Bradfield, C.A. The search for endogenous activators of the aryl hydrocarbon receptor. *Chem. Res. Toxicol.* **21**, 102-116 (2008).

45. Ikuta, T., Kobayashi, Y. & Kawajiri, K. Cell density regulates intracellular localization of aryl hydrocarbon receptor. *Journal of Biological Chemistry* **279**, 19209-19216 (2004).

46. Barhoover, M.A., Hall, J.M., Greenlee, W.F. & Thomas, R.S. Aryl Hydrocarbon Receptor Regulates Cell Cycle Progression in Human Breast Cancer Cells via a Functional Interaction with Cyclin-Dependent Kinase 4. *Molecular Pharmacology* **77**, 195-201 (2010).

47. Puga, A. et al. Aromatic hydrocarbon receptor interaction with the retinoblastoma protein potentiates repression of E2F-dependent transcription and cell cycle arrest. *Journal of Biological Chemistry* **275**, 2943-2950 (2000).

48. Tijet, N. et al. Aryl hydrocarbon receptor regulates distinct dioxin-dependent and dioxin-independent gene batteries. *Molecular Pharmacology* **69**, 140-153 (2006).

49. Uppstad, H., Øvrebø, S., Haugen, A. & Mollerup, S. Importance of CYP1A1 and CYP1B1 in bioactivation of benzo[a]pyrene in human lung cell lines. *Toxicol. Lett.* **192**, 221-228 (2010).

50. Pfeifer, G. et al. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene* **21**, 7435-7451 (2002).

51. Boysen, G. & Hecht, S. Analysis of DNA and protein adducts of benzo[alpha]pyrene in human tissues using structure-specific methods. *Mutat Res-Rev Mutat* **543**, 17-30 (2003).

52. Denissenko, M., Pao, A., Tang, M. & Pfeifer, G. Preferential formation of benzo a pyrene adducts at lung cancer mutational hotspots in P53. *Science* **274**, 430-432 (1996).

53. Gazdar, A. et al. Tumors of the Lung: Genetic and molecular alterations. *Pathology and Genetics of Tumours of the Lung, Pleura,Thymus and Heart* (2004).

54. DeMarini, D.M. et al. Lung tumor KRAS and TP53 mutations in nonsmokers reflect exposure to PAH-rich coal combustion emissions. *Cancer Res* **61**, 6679-6681 (2001).

55. Jirtle, R.L. & Skinner, M.K. Environmental epigenomics and disease susceptibility. *Nature Reviews Genetics* **8**, 253-262 (2007).

56. Wang, H. et al. Genetic susceptibility of lung cancer associated with common variants in the 3' untranslated regions of the adenosine triphosphate-binding cassette B1 (ABCB1) and ABCC1 candidate transporter genes for carcinogen export. *Cancer* **115**, 595-607 (2009).

57. Baccarelli, A. & Bollati, V. Epigenetics and environmental chemicals. *Curr. Opin. Pediatr.* **21**, 243-251 (2009).

58. Chuang, J.C. & Jones, P.A. Epigenetics and microRNAs. *Pediatr. Res.* **61**, 24R-29R (2007).

59. Jovanovic, J., Rønneberg, J.A., Tost, J. & Kristensen, V. The epigenetics of breast cancer. *Mol Oncol* **4**, 242-254 (2010).

60. Buschbeck, M. et al. The histone variant macroH2A is an epigenetic regulator of key developmental genes. *Nat Struct Mol Biol* **16**, 1074-1079 (2009).

61. Fire, A. et al. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-811 (1998).

62. Carthew, R.W. & Sontheimer, E.J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**, 642-655 (2009).

63. Lee, R.C., Feinbaum, R.L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-854 (1993).

64. Reinhart, B.J. et al. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**, 901-906 (2000).

65. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. & Enright, A.J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**, D140-4 (2006).

66. Monteys, A.M. et al. Structure and activity of putative intronic miRNA promoters. *RNA* **16**, 495-505 (2010).

67. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. Identification of novel genes coding for small expressed RNAs. *Science* **294**, 853-858 (2001).

68. Lee, Y., Jeon, K., Lee, J., Kim, S. & Kim, V.N. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J* **21**, 4663-4670 (2002).

69. Cai, X., Hagedorn, C.H. & Cullen, B.R. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* **10**, 1957-1966 (2004).

70. Lee, Y. et al. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* **23**, 4051-4060 (2004).

71. Zhou, X., Ruan, J., Wang, G. & Zhang, W. Characterization and identification of microRNA core promoters in four model species. *PLoS Comput. Biol.* **3**, e37 (2007).

72. Sernpere, L., Sokol, N., Dubrovsky, E., Berger, E. & Ambros, V. Temporal regulation of microRNA expression in Drosophila melanogaster mediated by hormonal signals and Broad-Complex gene activity. *Dev Biol* **259**, 9-18 (2003).

73. Johnson, S.M., Lin, S.Y. & Slack, F.J. The time of appearance of the C. elegans let-7 microRNA is transcriptionally controlled utilizing a temporal regulatory element in its promoter. *Dev Biol* **259**, 364-379 (2003).

74. Lee, Y. et al. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-419 (2003).

75. Filipowicz, W., Bhattacharyya, S. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature Reviews Genetics* **9**, 102-114 (2008).

76. Lund, E., Güttinger, S., Calado, A., Dahlberg, J.E. & Kutay, U. Nuclear export of microRNA precursors. *Science* **303**, 95-98 (2004).

77. Bernstein, E., Caudy, A.A., Hammond, S.M. & Hannon, G.J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363-366 (2001).

78. Hutvágner, G. et al. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, 834-838 (2001).

79. Cheloufi, S., Santos, dos, C., Chong, M. & Hannon, G. A Dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* **465**, 584-U76 (2010).

80. Hammond, S.M., Bernstein, E., Beach, D. & Hannon, G.J. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **404**, 293-296 (2000).

81. Lewis, B.P., Shih, I., Jones-Rhoades, M.W., Bartel, D.P. & Burge, C.B. Prediction of mammalian microRNA targets. *Cell* **115**, 787-798 (2003).

82. Tolia, N. & Joshua-Tor, L. Slicer and the Argonautes. *Nature Chemical Biology* **3**, 36-43 (2007).

83. Guo, H., Ingolia, N., Weissman, J. & Bartel, D. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835-840 (2010).

84. Jardim, M.J. microRNAs: Implications for air pollution research. *Mutat Res* (2011).doi: 10.1016/j.mrfmmm.2011.03.014

85. Baek, D. et al. The impact of microRNAs on protein output. *Nature* **455**, 64-71 (2008).

86. Enright, A.J. et al. MicroRNA targets in Drosophila. *Genome Biol* **5**, R1 (2003).

87. Bartel, D. MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136**, 215-233 (2009).

88. Calin, G.A. & Croce, C.M. MicroRNA signatures in human cancers. *Nature Reviews Cancer* **6**, 857-866 (2006).

89. Lynam-Lennon, N., Maher, S. & Reynolds, J. The roles of microRNA in cancer and apoptosis. *Biological Reviews* **84**, 55-71 (2009).

90. Martello, G. et al. A MicroRNA Targeting Dicer for Metastasis Control. *Cell* **141**, 1195-U176 (2010).

91. Karube, Y. et al. Reduced expression of Dicer associated with poor prognosis in lung cancer patients. *Cancer Science* **96**, 111-115 (2005).

92. Ventura, A. & Jacks, T. MicroRNAs and Cancer: Short RNAs Go a Long Way. *Cell* **136**, 586-591 (2009).

93. Du, L. et al. MicroRNA expression distinguishes SCLC from NSCLC lung tumor cells and suggests a possible pathological relationship between SCLCs and NSCLCs. *Journal of Experimental & Clinical Cancer Research* **29**, 75 (2010).

94. Schembri, F. et al. MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 2319-2324 (2009).

95. Izzotti, A. et al. Downregulation of microRNA expression in the lungs of rats exposed to cigarette smoke. *FASEB J.* **23**, 806-812 (2009).

96. Moffat, I.D. et al. microRNAs in adult rodent liver are refractory to dioxin treatment. *Toxicol. Sci.* **99**, 470-487 (2007).

97. Hudder, A. & Novak, R. miRNAs: Effectors of environmental influences on gene expression and disease. *Toxicological Sciences* **103**, 228-240 (2008).

98. Sagredo, C. et al. Quantitative analysis of benzo[a]pyrene biotransformation and adduct formation in Ahr knockout mice. *Toxicol. Lett.* **167**, 173-182 (2006).

99. Duan, H., Jiang, Y., Zhang, H. & Wu, Y. MiR-320 and miR-494 affect cell cycles of primary murine bronchial epithelial cells exposed to benzo[a]pyrene. *Toxicology in Vitro* **24**, 928-935 (2010).

100. Ramirez, R.D. et al. Immortalization of human bronchial epithelial cells in the absence of viral oncoproteins. *Cancer Res* **64**, 9027-9034 (2004).

101. Mullis, K.B. & Faloona, F.A. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Meth Enzymol* **155**, 335-350 (1987).

102. Kubista, M. et al. The real-time polymerase chain reaction. *Mol. Aspects Med.* **27**, 95-125 (2006).

103. Pfaffl, M.W. Relative quantification. *Real-time PCR* 63-82 (2006).

104. Souazé, F., Ntodou-Thomé, A., Tran, C.Y., Rostène, W. & Forgez, P. Quantitative RT-PCR: limits and accuracy. *BioTechniques* **21**, 280-285 (1996).

105. Shell, S. et al. Let-7 expression defines two differentiation stages of cancer. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 11400-11405 (2007).

106. Vandesompele, J. et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**, RESEARCH0034 (2002).

107. Mestdagh, P. et al. A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol* **10**, R64 (2009).

108. Choong, M.L., Yang, H.H. & McNiece, I. MicroRNA expression profiling during human cord blood-derived CD34 cell erythropoiesis. *Exp. Hematol.* **35**, 551-564 (2007).

109. Andersen, C.L., Jensen, J.L. & Ørntoft, T.F. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res* **64**, 5245-5250 (2004).

110. Motulsky, H. *Intuitive biostatistics*. 386 (Oxford University Press, USA: 1995).

111. Raychaudhuri, S., Stuart, J.M. & Altman, R.B. Principal components analysis to summarize microarray experiments: application to sporulation time series. *Pac Symp Biocomput* 455-466 (2000).

112. D'haeseleer, P. How does gene expression clustering work? *Nat Biotechnol* **23**, 1499-1501 (2005).

113. Lin, P., Yu, S. & Yang, P. MicroRNA in lung cancer. *Br. J. Cancer* **103**, 1144-1148 (2010).

114. Dorsam, R.T. & Gutkind, J.S. G-protein-coupled receptors and cancer. *Nature Reviews Cancer* **7**, 79-94 (2007).

115. Moody, T.W. et al. GRP receptors are present in non small cell lung cancer cells. *J. Cell. Biochem. Suppl.* **24**, 247-256 (1996).

116. Shriver, S.P. et al. Sex-specific expression of gastrin-releasing peptide receptor: relationship to smoking history and risk of lung cancer. *J Natl Cancer I* **92**, 24-33 (2000).

117. Miko, E. et al. DIFFERENTIALLY EXPRESSED MicroRNAs IN SMALL CELL LUNG CANCER. *Exp Lung Res* **35**, 646-664 (2009).

118. Liu, T., Tang, H., Lang, Y., Liu, M. & Li, X. MicroRNA-27a functions as an oncogene in gastric adenocarcinoma by targeting prohibitin. *Cancer Letters* **273**, 233-242 (2009).

119. Porkka, K.P. et al. MicroRNA Expression Profiling in Prostate Cancer. *Cancer Res* **67**, 6130-6135 (2007).

120. ZHU, H. et al. Role of MicroRNA miR-27a and miR-451 in the regulation of MDR1/P-glycoprotein expression in human cancer cells. *Biochemical Pharmacology* **76**, 582-588 (2008).

121. Mertens-Talcott, S.U., Chintharlapalli, S., Li, X. & Safe, S. The oncogenic microRNA-27a targets genes that regulate specificity protein transcription factors and the G2-M checkpoint in MDA-MB-231 breast cancer cells. *Cancer Res* **67**, 11001-11011 (2007).

122. Li, X. et al. MicroRNA-27a Indirectly Regulates Estrogen Receptor {alpha} Expression and Hormone Responsiveness in MCF-7 Breast Cancer Cells. *Endocrinology* **151**, 2462-2473 (2010).

123. Bhat-Nakshatri, P. et al. Estradiol-regulated microRNAs control estradiol response in breast cancer cells. *Nucleic Acids Res* **37**, 4850-4861 (2009).

124. Kiyohara, C., Yoshimasu, K., Shirakawa, T. & Hopkin, J.M. Genetic polymorphisms and environmental risk of lung cancer: a review. *Rev Environ Health* **19**, 15-38 (2004).

125. Singhal, R., Shankar, K., Badger, T.M. & Ronis, M.J. Estrogenic status modulates aryl hydrocarbon receptor--mediated hepatic gene expression and carcinogenicity. *Carcinogenesis* **29**, 227-236 (2008).

126. Ronchetti, D. et al. An integrative genomic approach reveals coordinated expression of intronic miR-335, miR-342, and miR-561 with deregulated host genes in multiple myeloma. *Bmc Med Genomics* **1**, - (2008).

127. Scarola, M., Schoeftner, S., Schneider, C. & Benetti, R. miR-335 directly targets Rb1 (pRb/p105) in a proximal connection to p53-dependent stress response. *Cancer Res* **70**, 6925-6933 (2010).

128. Najafi-Shoushtari, S.H. et al. MicroRNA-33 and the SREBP Host Genes Cooperate to Control Cholesterol Homeostasis. *Science* **328**, 1566-1569 (2010).

129. Yanaihara, N. et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* **9**, 189-198 (2006).

130. Yamazaki, H. et al. Glia maturation factor-beta is produced by thymoma and may promote intratumoral T-cell differentiation. *Histopathology* **47**, 292-302 (2005).

131. Gazdar, A.F., Gao, B. & Minna, J.D. Lung cancer cell lines: Useless artifacts or invaluable tools for medical science? *Lung Cancer* **68**, 309-318 (2010).

132. Pleasance, E.D. et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184-190 (2010).

133. Chung, J. et al. Abundance of aryl hydrocarbon receptor potentiates benzo[a]pyrene-induced apoptosis in Hepa1c1c7 cells via CYP1A1 activation. *Toxicology* **235**, 62-72 (2007).

134. Øvrevik, J. et al. Differential effects of nitro-PAHs and amino-PAHs on cytokine and chemokine responses in human bronchial epithelial BEAS-2B cells. *Toxicology and Applied Pharmacology* **242**, 270-280 (2010).

135. Soshilov, A. & Denison, M.S. Role of the Per/Arnt/Sim domains in ligand-dependent transformation of the aryl hydrocarbon receptor. *Journal of Biological Chemistry* **283**, 32995-33005 (2008).

136. Jardim, M.J., Fry, R.C., Jaspers, I., Dailey, L. & Diaz-Sanchez, D. Disruption of microRNA expression in human airway cells by diesel exhaust particles is linked to tumorigenesis-associated pathways. *Environ. Health Perspect.* **117**, 1745-1751 (2009).

Appendix

Content

A.Materials

Kits

	Kit name	Manufacturer (catalogue No.)
RNA quality	Agilent RNA 6000 Nano Kit	Agilent (5067-1511)
RT-qPCR		
mRNA cDNA synthesis	qScript™ cDNA Synthesis Kit	Quanta (95047-500)
mRNA qPCR	PerfeCTa SYBR Green Fast Mix, ROX™	Quanta (95073)
miRNA cDNA synthesis	miScript Reverse Transcription Kit	Qiagen (218061)
miRNA qPCR	miScript SYBR Green PCR Kit	Qiagen (218073)
RNA interference		**Invitrogen,**
Transfection	Lipofectamine™ RNAiMAX	13778100
Negative control	Stealth RNAi™ siRNA Hi GC	12935-400
Negative control	Stealth RNAi™ siRNA LO GC	12935-200
AHR siRNA	Stealth Select RNAi™ siRNA, Set of 3	1299003

PCR Primers

	Sequece	Maufacturer (catalogue No.)
Actin Up	5' GCGAGAAGATGACCCAGATCA	DNA Technology A/S
Actin Lo	5' GATAGCACAGCCTGGATAGCAA	DNA Technology A/S
AHR Up	5' CCCTTGGAAATTCATTGCCA	DNA Technology A/S
AHR Lo	5' GGAGAGGTGCTTCATATGTCGTC	DNA Technology A/S
CYP1A1 Up	5' CATCCCCCACAGCACAACA	DNA Technology A/S
CYP1A1 Lo	5' CAGGGGTGAGAAACCGTTCA	DNA Technology A/S
ARNT Up	5' TTCGTGAGCAGCTTTCCACTT	DNA Technology A/S
ARNT Lo	5' TCTCATGGAAGACTGCTGACCTT	DNA Technology A/S
AHRR Up	5' GAAGGGACGATATTTTATGCATCAG	DNA Technology A/S
AHRR Lo	5' CACGTGGATGTAGTCATAAATGTTCTG	DNA Technology A/S

	Sequece	Maufacturer (catalogue No.)
hsa-miR-27a*	5' UUCACAGUGGCUAAGUUCCGC	Qiagen (MS00003241)
hsa-miR-33a*	5' GUGCAUUGUAGUUGCAUUGCA	Qiagen (MS00003304)
hsa-miR-335*	5' UUUUUCAUUAUUGCUCCUGACC	Qiagen (MS00009464)
hsa-mIR-592	5' UUGUGUCAAUAUGCGAUGAUGU	Qiagen (MS00004914)
hsa-miR-122a	5' UGGAGUGUGACAAUGGUGUUUG	Qiagen (MS00003416)
hsa-miR-203	5' GUGAAAUGUUUAGGACCACUAG	Qiagen (MS00003766)
hsa-miR-494	5'UGAAACAUACACGGGAAACCUC	Qiagen (MS00004340)

Instruments

Cell Counter	Countess automated cell counter, Invitrogen	Spectrophotometer	Eppendorf Biophotometer
Centrifuges	Eppendorf Centrifuge 5702	Microscope	Nikon Diaphoto light microscope
	Sigma 2-6E	Multimode reader	Modulus Microplate Reader, Pronova
Cooler centrifuges	Eppendorf Centrifuge 5417R	qPCR instrument	ABI Prism 7900 HT Sequence Detection System, Applied Biosystems
	Sigma 4K15	Thermal cycler	Perkin Elmer Cetus DNA Cycler 480
Heating block	Grand instruments QBT2	Pipette robot	Eppendorf EP-Motion

Data software

PASW Statistics 18.0 (IBM SPSS, Somers, NY, USA)
SDS Software 2.4 (Applied Biosystems, Carlsbad, CA, USA)
Mathematic 8 (Wolfram Research, Champaign, IL, USA)
J-Express 2009 (MolMine A/S, Bergen, Norway)
Microsoft Office 2011 Excel (Microsoft Corporation, Redmond, WA)
MetaCore (GeneGo, Thompson Reuters, St. Joseph, MI, USA)

Cell culture media

GIBCO® Opti-MEM® I Reduced-Serum Medium	HEPES buffer, 2,400 mg/L sodium bicarbonate, hypoxanthine, thymidine, sodium pyruvate, L-glutamine, trace elements, and growth factors, no phenol red.	Invitrogen	11058-021
GIBCO® D-MEM/F-12	Contains 15 mM HEPES buffer, and L-glutamine, no phenol red	Invitrogen	11039-021
GIBCO® LHC-9 Serum-free medium	L-glutamine, Serum-free, phenol red	Invitrogen	12680-013

Chemicals

Acetic Acid (glacial)	Merck	KCl	Merck
CellTiter-Blue® Cell Viability Assay	Promega	Leibovitz's L15 media	Invitrogen
Chloroform	Sigma Aldrich	$MgSO_4$	Sigma Aldrich
Collagen		NaCl	Merck
Dimethylsulfoxid	Sigma Aldrich	NaOH	Merck
DNase I (bovine pancreas)	Sigma Aldrich	$NaHPO_4$* H_2O	Merck
EDTA (Triplex III)	Merck	N,N,N',N'-tetra-methyl-ethylendiamin (TEMED)	
Ethanol (rectified, absolute)	Kemethyl	Penicillin Streptomycin	Gibco
Fetal Bovine Serum (FBS)	Gibco	Phenol red	Sigma Aldrich
GelRed	Biotium Inc.	Propane-2-ol (isopropanol)	Merck
Glyserol		Thapsigarin	Sigma Aldrich
HEPES	Sigma Aldrich	Tris Base 7-9®	Sigma Aldrich
HCl	Merck	Trypan Blue Stain	Invitrogen
TRIzol RNA lysis reagent	5' prime	Trypsin	Sigma Aldrich

Solutions

All solutions are made with ddH_2O and sterile filtered before use.

AF-media
76% L-15 medium, 2% 1M HEPES, 2% PS, 20% FBS

BSA 60 mg/ml stock
60 mg BSA, 1 ml PBS

Collagen
1% collagen solution, 3.13 mg/ml, 99% HBS

DEPC water
0.1% DEPC in ddH_2O

DMSO for cell culture storage
50% L-15 medium, 2% 1M HEPES, 8% DMSO, 40% FBS

DNase (4 mg/ml)
4 g DNase I, 87.66 mg NaCl, 5.06 mg $MgSO_4$, H_2O to 1 ml

Ethanol in DEPC water (75%)
12.5 ml DEPC H_2O, Absolute ethanol to 50 ml

EDTA
9.3 EDTA, 50 ml, H_2O, pH adjusted to 8.

Fetal Bovine Serum (FBS)
Heat inactivated at 56°C for 45 min.

HEPES
238.3 g HEPES, 1 ml 0.12% phenol red in 1 L ddH_2O, pH 7.3

Phenol red
125 g phenol red, 360 µl 1M NaOH, 100 ml H_2O

Phosphate Buffered Saline (PBS)
7.07 g NaCl, 0.20 g KCl, 1.94 g $NaHPO_4$*H_2O, H_2O to 1 L.

Tris-EDTA (TE) Buffer
500 µl Tris-Cl pH8, 10 µl EDTA, H_2O to 50 ml.

Tris-acetate-EDTA (TAE) buffer
242 g Tris base, 57.1 ml glacial acetic acid, 100 ml 0.5 M EDTA pH8.

Tris-Cl 1M pH8
121.1 g Tris base, 42 ml HCl, H_2O to 1 L.

Trypsin 1%
50 mg trypsin, 50 ml PBS.

B. General cell culture work

B.1. Thawing cells

1. An ampoule was taken from the nitrogen tank and thawed in a water bath (37°C).
2. The cells were transferred to a centrifuge tube.
3. PBS (5 mL) was added and the tube was centrifuged (1000 rpm, 4 minutes).
4. The supernatant was removed and the pellet was re-suspended in medium (3 ml).
5. The cell suspension was pipetted into a petri dish (100 mm) already containing 5 mL of medium (8 ml in total).
6. The petri dish was incubated at 37°C.
7. Passaging cells (100 mm dishes)
8. The media over the cells was removed.
9. The petridish was washed twice with PBS (2 x 10 ml).
10. Trypsin (1 ml, PRODUSENT) was added to the dish, and the cells were incubated (37°C) until they had loosened from the dish. This was checked in the microscope.
11. LHC-9 medium (5 ml) was added to the dish and the cell suspension was pipetted to a centrifuge tube and centrifuged (1000 rpm, 4 minutes).
12. The supernatant was removed.
13. The pellet was re-suspended in medium. Amount of medium used depended on the desired dilution of the cells.
14. Cell suspension (1 ml) was pipetted to a new petri dish added 7 ml LHC-9 medium (8 mL in total pr. 100 mm dish).
15. Cells were incubated at 37°C.

B.2.Freezing cells

Same procedure as passaging of cells is followed to point number 5

1. The cell pellet was re-suspended in AF (500 µL) and transferred to a freezing vial.
2. DMSO freezing solution (8%, 500 µl) was added and the suspension was mixed well.
3. The vial was put in an insulated box and placed at -80°C, 4-6 hours, before placed in liquid nitrogen for long term storage.

B.3.Collagen coating of petri dishes or cell culture plates

Purified collagen solution for cell culture (0.012 M HCl, 2.7 mg/ml) was diluted in HBS to a total concentration of 0.03 mg/ml. Appropriate volume of collagen was added to the petridishes or plates, so that it was completely covered. The dishes/plates were incubated at room temperature in the LAF bench for 2 hours. The collagen solution was removed and the dishes were washed with PBS.

B.4. Cell culture maintenance

HBEC3-KT cells were maintained in petri dishes (100 mm) until seeded for experiments. The cells were given fresh LHC-9 medium three times a week and passaged when reaching ~80% confluence.

C.Pre-processing of miRNA expression raw data

The RT-qPCR raw data was normalised by using normalisation factors (NF) calculated in GeNorm (Table B.1). GeNorm use expression quantities to calculate an M-value describing the stability of each gene relative to all other genes. The gene with the lowest M-value is the most stable (Figure B.1). The NF was calculated by taking the geometric mean of the most stably expressed miRNAs. Expression values of raw C_q was stabilised by dividing the gene quantity of each sample on the appropriate NF.

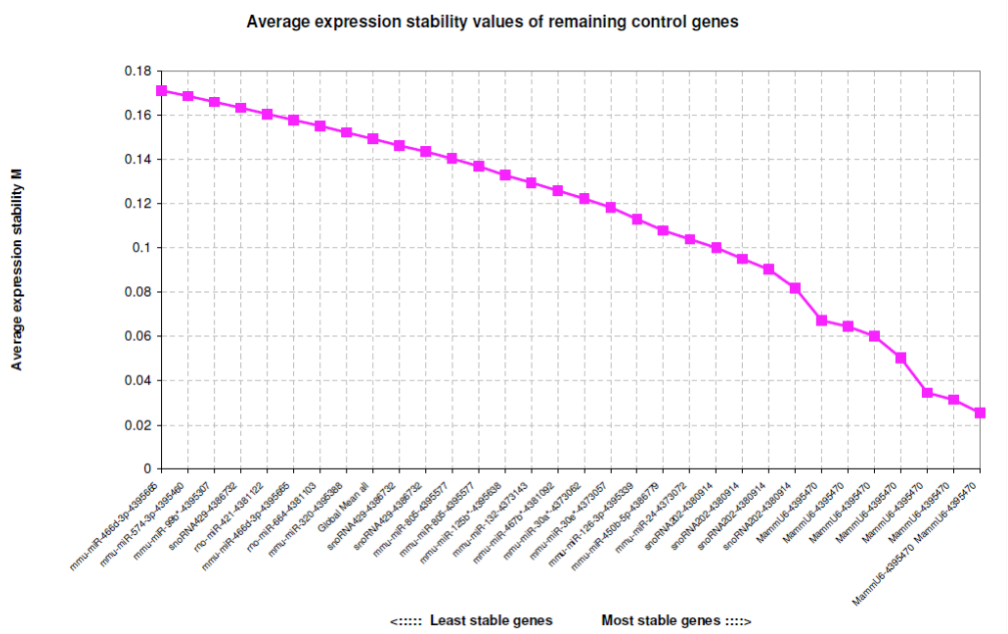


Figure B.1: Stability of microRNAs across all samples. The 32 genes with lowest variance (calculated in Excel) was imported to GeNorm. In addition, the mean of the global miRNA expression level for each sample were imported. Some genes occurred in multiple replicates on the LDA cards. Stability increases with decreasing M, indicating the mammalian RNU6 genes as the most stably expressed.

Table B.1: Normalization factor (NF) calculated in GeNorm by taking the geometric mean of the quantities of the most stably expressed microRNAs.

Mouse	NF	Genotype	Sex
LU212	0.9550	Ahr+/+	Female
LU216	1.0010	Ahr+/+	Female
LU227	1.1254	Ahr+/+	Female
LU236	1.0767	Ahr+/+	Male
LU244	0.8893	Ahr+/+	Male
LU255	0.8873	Ahr+/+	Male
LU201	0.9438	Ahr-/-	Female
LU211	1.0160	Ahr-/-	Female
LU215	0.9840	Ahr-/-	Female
LU232	1.1036	Ahr-/-	Male
LU234	1.0592	Ahr-/-	Male
LU246	0.9919	Ahr-/-	Male

D.Sequence of miRNAs assayed in human cell line

Expression of seven miRNAs, four from the mouse study in this thesis and three selected from the literature, were investigated in immortalised human lung cell lines. Sequence of mouse (mmu) and human (hsa) miRNAs and their genomic location is shown in Table C.1.

Table C.1: Sequence of mouse (mmu) and human (hsa) miRNAs and their genomic locations.

miRNA name	Sequence	Location	miRNA name	Sequence	Location
mmu-miR-27a	UUCACAGUGGCUAAGUUCCGC	Chr. 8 Cluster: *mir-23a, mir-27a, mir-24-2, mir-3074-2*	hsa-miR-27a	UUCACAGUGGCUAAGUUCCGC	Chr. 19 Cluster: *mir-23a, mir-27a, mir-24-2*
mmu-miR-27a*	AGGGCUUAGCUGCUUGUGAGCA		hsa-miR-27a*	AGGGCUUAGCUGCUUGUGAGCA	
mmu-miR-33	GUGCAUUGUAGUUGCAUUGCA	Chr. 15 Host gene: Srebf2	hsa-miR-33a	GUGCAUUGUAGUUGCAUUGCA	Chr. 22 Host gene: SREBF2
mmu-miR-33*	CAAUGUUUCCACAGUGCAUCAC		hsa-miR-33a*	CAAUGUUUCCACAGUGCAUCAC	
mmu-miR-335-5p	UCAAGAGCAAUAACGAAAAAUGU	Chr. 6 Host gene: Mest	hsa-miR-335	UCAAGAGCAAUAACGAAAAAUGU	Chr. 7 Host gene: MEST
mmu-miR-335-3p	UUUUUCAUUAUUGCUCCUGACC		hsa-miR-335*	UUUUUCAUUAUUGCUCCUGACC	
mmu-miR-592	AUUGUGUCAAUAUGCGAUGAUGU	Chr. 6 Host gene: Grm8	hsa-miR-592	UUGUGUCAAUAUGCGAUGAUGU	Chr. 7 Host gene: GRM8
mmu-miR-592*	UCAUCACGUGGUGACGCAACAU		No antisense sequence reported		

miRNA name	Sequence	Location	miRNA name	Sequence	Location
mmu-miR-122a	UGGAGUGUGACAAUGGUGUUUG	Chr. 8	hsa-miR-122a	UGGAGUGUGACAAUGGUGUUUG	Chr. 18 Cluster: mir-3591
mmu-miR-122a*	AAACGCCAUUAUCACACUAA		hsa-miR-122a*	AACGCCAUUAUCACACUAAAUA	
mmu-miR-203	GUGAAAUGUUUAGGACCACUAG	Chr. 12	hsa-miR-203	GUGAAAUGUUUAGGACCACUAG	Chr. 14 Cluster: mir-3545
mmu-miR-203*	AGUGGUUCUUGACAGUUCAACA		No antisense sequence reported		
mmu-miR-494	UGAAACAUACACGGGAAACCUC	Chr. 12 Cluster: mir-379, mir-411, mir-299, mir-380, mir-1197, mir-323, mir-758, mir-329, mir-494, mir-679, mir-1193, mir-666, mir-543, mir-495, mir-667, mir-376c, mir-654, mir-376b, mir-376a, mir-300	hsa-miR-494	UGAAACAUACACGGGAAACCUC	Chr. 14 Cluster: mir-379, mir-411, mir-299, mir-380, mir-1197, mir-323, mir-758, mir-329-1, mir-329-2, mir-494, mir-1193, mir-543, mir-495
mmu-miR-494*	AGGUUGUCCGUGUUGUCUUCUC		No antisense sequence reported		

E. Genome database search for AHRE

A search for AHR response elements (AHRE) in the promoter region of the four most differentially expressed microRNAs (miR-592, miR-33*, miR-335-3p, and miR-27a*) was performed in mathematica 8 (Wolfram Research) in both mouse and human genome. The human genome is implemented in the Mathematica Software, while the mouse genome (version 37) was downloaded from the National Center for Biotechnology Information (NCBI) ftp server (available from ftp.ncbi.nlm.nih.gov) an imported to Mathematica. The search was limited to 10 kb upstream and 1 kb downstream of the pre-miRNA sequence. These ranges were chosen based on similar type of search previously reported in the literature[96]. Search for both full and core sequence of AHRE-I, and of full AHRE-II was performed. miRNA sequences where downloaded from miRBase (http://mirbase.org)

Defining AHRE

```
ahreI=RegularExpression["[TG][AGCT]GCGTG[AC][GC]A"];
ahreII=RegularExpression["CATG[ATGC][ATGC][ATGC][ATGC][ATGC][ATGC]C[TA]TG"];
```

Defining human pre-miRNA sequence

```
fullHSAmir27a="CUGAGGAGCAGGGCUUAGCUGCUUGUGAGCAGGGUCCACACCAAG
UCGUGUUCACAGUGGCUAAGUCCGCCCCCCAG"
```

Search for pre-miRNA sequence in human genome

```
GenomeLookup[fullHSAmir27a]
```

Extracting sequence -10 kb to +1 kb of pre-miRNA sequence

```
mir27aflanking = GenomeData[{{"Chromosome19", -1}, {50003321 - 10000, 50003398 +
1000}}]
```

Search for AHRE-I and AHRE-II in -10 kb to +1 kb extracted sequence

```
StringPosition[mir27aflanking, ahreI]
StringPosition[mir27aflanking, ahreII]
```

Defining mouse pre-miRNA sequence

Firma

fullMMUmir27a =
"UGGCCUGAGGAGCAGGGCUUAGCUGCUUGUGAGCAAGGUCCACAGCAAAGUCGU
GUUCACAGUGGCUAGUUCCGCCCCCUGGACCC";

Search for pre-miRNA sequence in mouse genome
StringPosition[mmchr8, StringReplace[fullMMUmir27a, "U" -> "T"]]

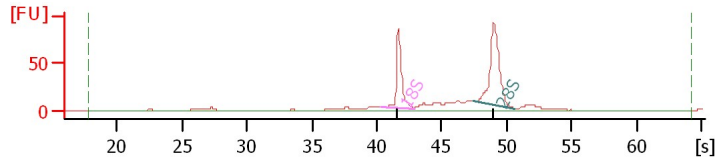
Extracting sequence -10 kb to +1 kb of pre-miRNA sequence
mmumiR27aflanking = StringTake[mmchr8[[6]], {11565925 - 10000, 11566011 + 1000}]

Search for AHRE-I and AHRE-II in -10 kb to +1 kb extracted secuense
StringPosition[mmumiR27aflanking, ahreI]
StringPosition[mmumiR27aflanking, ahreII]

F.Analysis of RNA quality

The quality of RNA isolated by TRIzol was assayed on a subset of samples by using Agilent RNA 6000 Nano Kit and Agilent 2100 bioanalyzer (Agilent Technologies). The chips contains interconnected microchannels that is used for separation of nucleic acid fragments based on their size as they are driven through it electrophoretically. The procedure were performed as described in protocol provided by producer. The quality of the samples are given as a RIN value from 1-10 where 10 indicates theoretically completely pure and intact RNA. RIN values > 5 is acceptable.

A HBEC3-KT, B[*a*]P exposure.



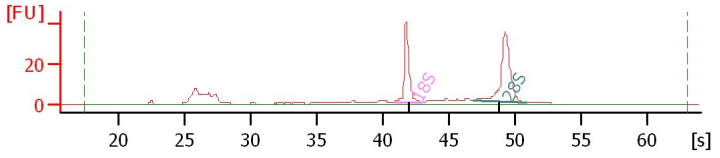
Overall Results for sample 2 : 21

RNA Area:	434.7
RNA Concentration:	4,030 ng/µl
rRNA Ratio [28s / 18s]:	1.7
RNA Integrity Number (RIN):	9.4 (B.02.02)

Fragment table for sample 2 : 21

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	40.34	42.86	87.4	20.1
28S	47.38	50.59	147.5	33.9

B HBEC3-KT, B[*a*]P exposure, control siRNA.



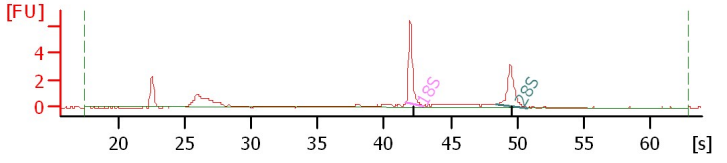
Overall Results for sample 4 : 221

RNA Area:	221.3
RNA Concentration:	2,052 ng/µl
rRNA Ratio [28s / 18s]:	1.5
RNA Integrity Number (RIN):	8.7 (B.02.02)

Fragment table for sample 4 : 221

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	40.88	43.03	37.3	16.8
28S	46.88	50.95	54.6	24.7

C HBEC2-KT, B[*a*]P exposure



Overall Results for sample 5 : 244

RNA Area:	21.2
RNA Concentration:	197 ng/µl
rRNA Ratio [28s / 18s]:	0.8
RNA Integrity Number (RIN):	8.4 (B.02.02)

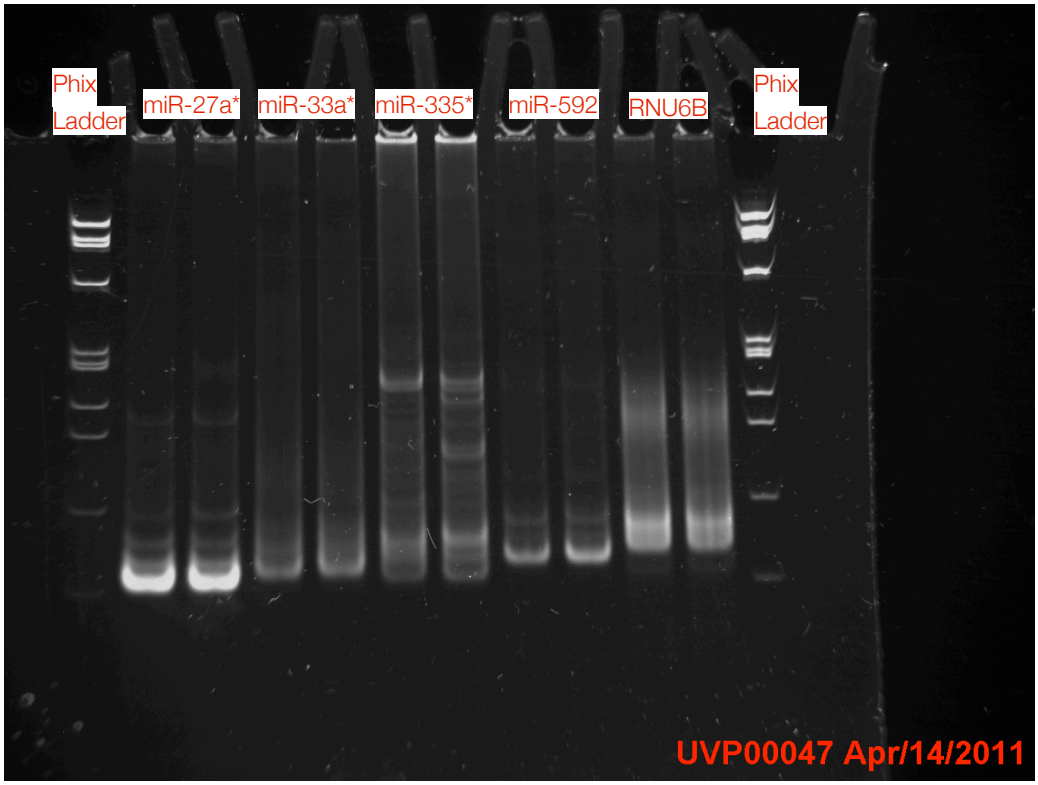
Fragment table for sample 5 : 244

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	41.67	42.90	4.7	22.1
28S	48.38	50.80	3.7	17.6

Figure E.1: Electrophorograms of three RNA samples. RIN values are: A: 9.4, B: 8.7, C: 8.4.

G.Poly acrylamide gel electrophoresis (PAGE)

Poly acrylamide gel electrophoresis can be used to separate DNA fragments under 500 nt of length. The method is very specific and can be used to separate DNA molecules varying by only one nucleotide. PAGE was used to analyse PCR products from the RT-qPCR of miRNAs. Since the miRNA PCR products are very small a gel concentration of 10% was used.



H. MetaCore symbols