DISSERTATIONES
BIOLOGICAE
UNIVERSITATIS
TARTUENSIS
293

URMO VÕSA

MicroRNAs in disease and health: aberrant regulation in lung cancer and association with genomic variation

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
293

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
293

URMO VÕSA

MicroRNAs in disease and health: aberrant regulation in lung cancer and association with genomic variation

UNIVERSITY OF TARTU
Press

Institute of Molecular and Cell Biology, University of Tartu, Estonia

This dissertation was accepted for the commencement of the degree of Doctor of Philosophy (in Molecular and Cell Biology) on May 13, 2016 by the Council of the Institute of Molecular and Cell Biology at the University of Tartu.

Supervisor: Tarmo Annilo, PhD, Senior Reseacher
Estonian Genome Center
University of Tartu
Tartu, Estonia

Reviewer: Prof. Maido Remm, PhD
Chair of Bioinformatics
Institute of Molecular and Cell Biology
University of Tartu
Tartu, Estonia

Opponent: Manlio Vinciguerra, PhD, Principal Research Associate
Institute for Liver and Digestive Health
Division of Medicine
Faculty of Medical Sciences
University College London
London, United Kingdom

Commencement: Room No. 105, 23B Riia St., Tartu, on August 23, 2016, at 10.15 am.

The publication of this dissertation is granted by Institute of Molecular and Cell Biology at the University of Tartu.

ISSN 1024-6479
ISBN 978-9949-77-135-6 (print)
ISBN 978-9949-77-136-3 (pdf)

University of Tartu Press
www.tyk.ee

“Now I'm a scientific expert;
that means I know nothing about absolutely everything.”

Arthur C. Clarke

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

Ref I Võsa U, Vooder T, Kolde R, Fischer K, Välk K, et al. 2011. Identification of miR-374a as a prognostic marker for survival in patients with early-stage non-small cell lung cancer. *Genes, Chromosom. Cancer*. 50(10):812–22

Ref II Võsa U, Vooder T, Kolde R, Vilo J, Metspalu A, Annilo T. 2013. Meta-analysis of microRNA expression in lung cancer. *Int. J. Cancer*. 132(12):2884–93

Ref III Võsa U, Esko T, Kasela S, Annilo T. 2015. Altered gene expression associated with microRNA binding site polymorphisms. *PLoS One*. 10(10):e0141351

The publications listed above have been reprinted with the permission of the copyright owners.

My contributions to the above publications are as follows:

Ref I Participated in conducting the experiments, analyzed the data, and participated in writing the manuscript.

Ref II Mined the data, performed the analyses, and drafted the manuscript.

Ref III Participated in the study design, mined the data, performed the analyses, and drafted the manuscript.

ABBREVIATIONS

AD	Adenocarcinoma
AMO	Anti-miRNA oligonucleotide
AntagomiR	Specific subclass of anti-microRNA oligonucleotides with attached cholesterol group in the 3' end
BA	Bronchioalveolar carcinoma
BCL	B-cell lymphoma
CDS	Coding DNA sequence
CLL	Chronic lymphocytic leukemia
eSNP	Expression affecting SNP
FC	Fold change
GWAS	Genome-wide association study
IsomiR	miRNA sequence differing from canonical sequence encoded by the same locus
LCC	Large cell carcinoma
LNA	Locked nucleic acid
miRNA	MicroRNA
miRNome	All microRNAs expressed in the specific organism/tissue
Mirtron	MicroRNA transcribed from introns independently of microprocessor complex by spliceosome
MoR	miRNA-offset RNA
MRE	MicroRNA recognition element
MRE-SNP	Single nucleotide polymorphism located in microRNA recognition element
NSCLC	Non-small cell lung cancer
OncomiR	MicroRNA with oncogenic function
ORF	Open reading frame
RBP	RNA-binding protein
SCC	Squamous cell carcinoma
Simtron	MicroRNA transcribed from introns independently of microprocessor complex and spliceosome
siRNA	Small interfering RNA
SNP	Single-nucleotide polymorphism
Targetome	Target genes of one or multiple miRNAs

INTRODUCTION

Until recently, the protein-coding region of the human genome was considered to be the most relevant portion, as it encodes the functional building blocks of the organism. However, the protein-coding region represents only a small part of the human genome (1~2%, consisting of exons of ~20,000 protein-coding genes). Recent progress in high-throughput sequencing technology and the concurrent "noncoding revolution" have demonstrated that the number of noncoding genes in humans is much higher than previously thought, even exceeding the number of protein-coding genes (GENCODE v23, 2015). Many noncoding transcripts are functional and crucial for normal functioning of the organism.

One of the best-studied subclasses of noncoding RNAs, microRNAs (miRNAs) are small single-stranded noncoding RNA molecules that regulate gene expression in a posttranscriptional manner. During the last 15 years, miRNAs have been in the limelight of human genetics due to their involvement in the development of several pathologies, particularly cancers. Because their expression profiles are altered in malignancies, miRNAs have been investigated as potential drug targets, and miRNAs with oncogenic or tumor-suppressor activities have been identified. The first fruits of this new knowledge are now becoming available for general public. At the time of writing, the first miRNA-based cancer drug, MRX34, is being studied in Phase I trials and several other drugs are expected to follow.

Expression profiles of several miRNAs have been observed to correlate with cancer (sub)type, stage, and patient prognosis. Therefore, these miRNAs may potentially serve as biomarkers. As miRNAs are stable in different body fluids, they have the potential to act as valuable noninvasive biomarkers, which could greatly simplify the treatment decisions of physicians in a safe and patient-friendly way. Such knowledge is already entering the clinic, as several diagnostic miRNA-based assays have become commercially available in recent years.

As with any complex trait, cancer predisposition involves the combinatorial effect of several risk factors. Over the last 10 years, intensive genome-wide association studies have identified many cancer-related genetic variants. However, mechanisms for how these variants translate into disease predisposition or phenotype remain poorly understood. Recent studies have integrated an additional layer of information in the analyses, investigating the influence of genetic variants on the global transcriptome. One potential mechanism for how genetic variants exert their effects on the transcriptome is through miRNA regulatory networks.

In this thesis, I give an overview of the biology of miRNAs, their relevance in cancer, and their interaction with genetic variation. In the experimental part, I describe the investigation of miRNA expression changes in a cohort of Estonian patients with non-small cell lung cancer, identifying miRNAs that may serve as

potential drug targets or biomarkers. Integrating the results of this experimental study with previous information, a systematic meta-analysis of differentially expressed miRNAs was performed, and a robust set of miRNAs that are up- or downregulated in non-small cell lung cancer was identified. Finally, it was investigated how genetic variation can influence miRNA activity, whether this interaction is reflected in the variation of the transcriptome, and whether miRNA-mediated mechanisms could contribute to the formation of phenotype or disease risk.

1. LITERATURE REVIEW

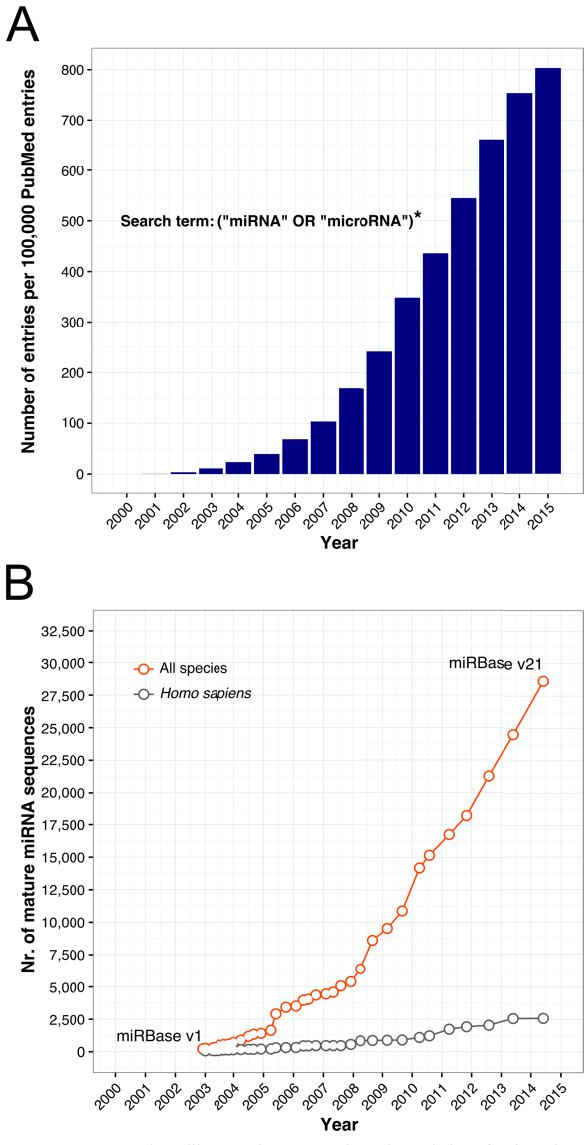
1.1. MicroRNAs

1.1.1. Overview

The history of microRNA (miRNA) research dates back to the year 1993, when Victor Ambros, Gary Ruvkun, and their coworkers investigated the regulation of *lin-14* translation by *lin-4* in *Caenorhabditis elegans.* Strikingly, they discovered that *lin-4* does not encode a protein, but rather encodes two small (~22- and ~61-nucleotide [nt]-long) transcripts (Lee et al., 1993). In-depth analyses revealed that the short RNA encoded by *lin-4* is partially complementary to the 3' untranslated region (3' UTR) of *lin-14* and downregulates the transcript by antisense interaction (Lee et al., 1993; Wightman et al., 1993). At the time, this mechanism was believed to be an isolated case, rather than a widespread phenomenon. The second important milestone in the field was the discovery of RNA interference (RNAi) (Fire et al., 1998) and its functionality in plants via small noncoding interfering RNAs (siRNAs) (Hamilton and Baulcombe, 1999).

Support for the idea that miRNAs might represent a universal mechanism of gene regulation came with the discovery of another small noncoding RNA in *C. elegans*. Specifically, *let-7* was shown to regulate the number of developmental genes in a time-dependent manner and, similarly to *lin-4*, to play an important role in developmental timing (Reinhart et al., 2000). As this RNA was found to be conserved in several eukaryotes (Pasquinelli et al., 2000), its functional relevance was highly likely. Subsequent RNA cloning studies identified hundreds of miRNAs in worm, fly, and human (Lagos-Quintana, 2001; Lau, 2001; Lee, 2001). With these discoveries, the miRNA field moved into the spotlight and during following years, the number of miRNA-related publications rose sharply (Fig. 1A).

To date, miRNAs have been identified in animals (Lagos-Quintana, 2001; Lee et al., 1993), plants (Reinhart, 2002) and some protists (Lin et al., 2009). There are some differences in the miRNA biogenesis pathways and in how miRNAs regulate their targets in plants compared to animals. As this thesis focuses on mammalian (human) miRNAs, I do not provide an extensive discussion of miRNA counterparts in plants. Researchers have described a class of miRNA-like RNAs (milRNAs) in fungi, the biogenesis pathways of which are considerably different from those of miRNAs in other organisms (Lee et al., 2010). Several viruses have been shown to encode miRNAs that are capable of regulating the expression of both viral and host genes (Bai et al., 2008). Positive correlation between miRNome size and organism complexity has led to the hypothesis that the acquisition of miRNA-mediated gene regulation allowed the emergence of highly complex organisms (Heimberg et al., 2008). In animals, miRNAs have been implicated in a wide range of biological processes, ranging from development (Reinhart et al., 2000) to defense against pathogens (Ma et al., 2011a) and even to the formation of memory (Gao et al., 2010a).



*Medline trend: automated yearly statistics of PubMed results for any query, 2004. Web resource at URL:http://dan.corlan.net/medline-trend.html. Accessed: 2016-04-09. (Archived by WebCite at http://www.webcitation.org/65RkD48SV)

Figure 1. Increasing knowledge about miRNAs. A. Yearly number of miRNA-related publications in PubMed. B. Number of known mature miRNAs in the database miRBase.

The reference database miRBase collects and curates identified miRNAs (Griffiths-Jones, 2004, 2006). Its latest version (v21) contains 28,645 mature

miRNA sequences from 233 species, including 2,588 human sequences. Although researchers are required to provide credible evidence about the validity of discovered miRNA(s), the database likely contains a number of incorrectly annotated sequences (Hansen et al., 2011). The number of newly discovered mature miRNAs continues to increase, albeit at a slower pace for human miRNAs, probably due to extensive prior research (Fig. 1B).

Nomenclature of miRNA transcripts follows specific rules. The name indicates the species (e.g., hsa- for human), RNA class (miRNA precursor: mir, mature miRNA: miR), and a number representing the order of discovery (Fig. 2). If identical mature miRNA sequences are encoded from multiple genomic loci, then the corresponding precursor sequences and mature miRNAs are denoted with additional numbers. If a group of miRNAs has high sequence similarity on the mature-sequence level, then the miRNAs are grouped into families, members of which are distinguished by letters (e.g., hsa-mir-200a and hsa-mir-200b). A suffix is used to indicate which arm of the miRNA precursor sequence the mature miRNA originates from (e.g., hsa-miR-205-5p from the 5' arm and hsa-miR-205-3p from 3' arm of the precursor). Older nomenclature distinguished minor nonfunctional "passenger" strand miRNAs as "star" sequences (e.g., hsa-miR-205*); however, this nomenclature was replaced by the 3'-5' system for clarity and because it is often difficult to determine the predominant arm.

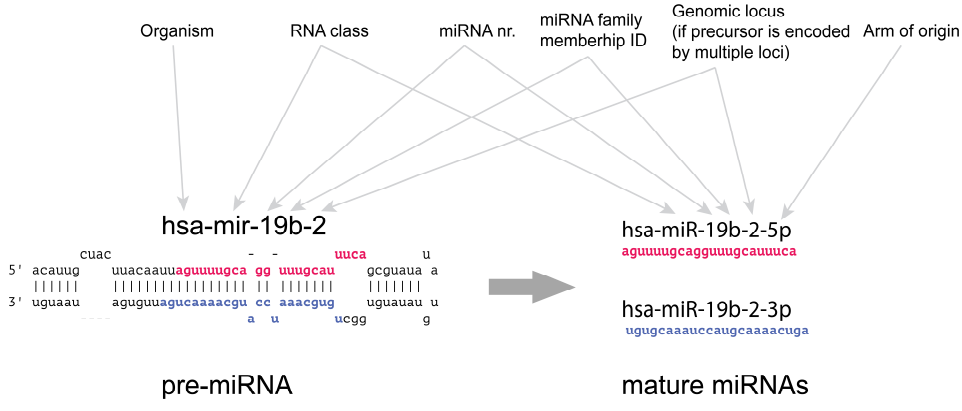


Figure 2. Nomenclature of miRNAs.

Names of some of the earliest discovered miRNAs (e.g., hsa-let-7 family) deviate from this nomenclature for historical reasons. As miRBase has updated the names of previously existing miRNAs several times, all miRNA names in this thesis are standardized to match miRBase v21. Additionally, species indicators are omitted for the sake of brevity.

1.1.2. Genomic organization, biogenesis, and modification of miRNAs

MicroRNAs are encoded by their own genes (intergenic miRNAs) or by transcriptional units residing in the introns or exons of protein-coding host genes (intragenic miRNAs) (Ambros et al., 2003; Rodriguez, 2004). Intron-encoded miRNAs may be transcribed together with the host gene (Baskerville and Bartel, 2005) or independently by their own promoter. According to a recent annotation of known and novel miRNAs, most of the known miRNAs map to intergenic regions (~70%) and introns (~12%). Smaller proportions map to repeat regions (~8%), exons (~5%), and long noncoding transcripts (~5%) (Londin et al., 2015). In human, there are almost equal numbers of known intergenic and intragenic miRNA genes (Paczynska et al., 2015). Many miRNAs are conserved in metazoans (Pasquinelli et al., 2000), suggesting their involvement in fundamental biological processes.

In animals, canonical miRNAs are transcribed by RNA polymerase II (Lee et al., 2004) as polyadenylated and capped primary transcripts (pri-miRNAs). MicroRNAs may be transcribed alone or in clusters, in which one pri-miRNA contains precursor hairpins for several miRNAs. The pri-miRNA is cut by the microprocessor complex, which consists of two major components: the endonuclease III Drosha (Lee et al., 2003) cuts RNA, and Di George critical region 8 (DGCR8) binds the double-stranded RNA (dsRNA), guiding Drosha into the correct site (Han et al., 2004). The resulting ~70-nt-long precursor sequence (pre-miRNA) consists of one hairpin structure, a 5'-phosphate group, and a 2-nt overhang on the 3' end. Pre-miRNA binds with a complex of exportin 5 (Exp 5) and Ran-GTP, which transports the complex to the cell cytoplasm (Lund and Gu, 2004; Yi et al., 2003).

The next step of preprocessing is another cleavage by the endonuclease III class enzyme Dicer (Bernstein et al., 2001; Grishok et al., 2001; Hutvágner et al., 2001; Ketting et al., 2001). Dicer forms a complex with several cofactors, including TAR DNA-binding protein (TARDBP), transactivating response RNA-binding protein (TRBP) (Chendrimada et al., 2005), and/or protein activator of interferon-induced protein kinase (PACT) (Lee et al., 2006), which mediate the substrate specificity of Dicer. The resulting ~21-nt-long duplex is loaded into the RNA-induced silencing complex (RISC), which consists of Argonaute proteins (in human, AGO1–4) and GW182. In human, AGO2 is the only member of the Argonaute family capable of cleaving the target strand (Höck and Meister, 2008). Other members of the protein family function as structural components of RISC or participate in translational repression of the target gene. GW182 acts as a "bridge" allowing RISC to interact with additional protein complexes involved in the downregulation of gene activity (Behm-Ansmant et al., 2006; Rehwinkel et al., 2005).

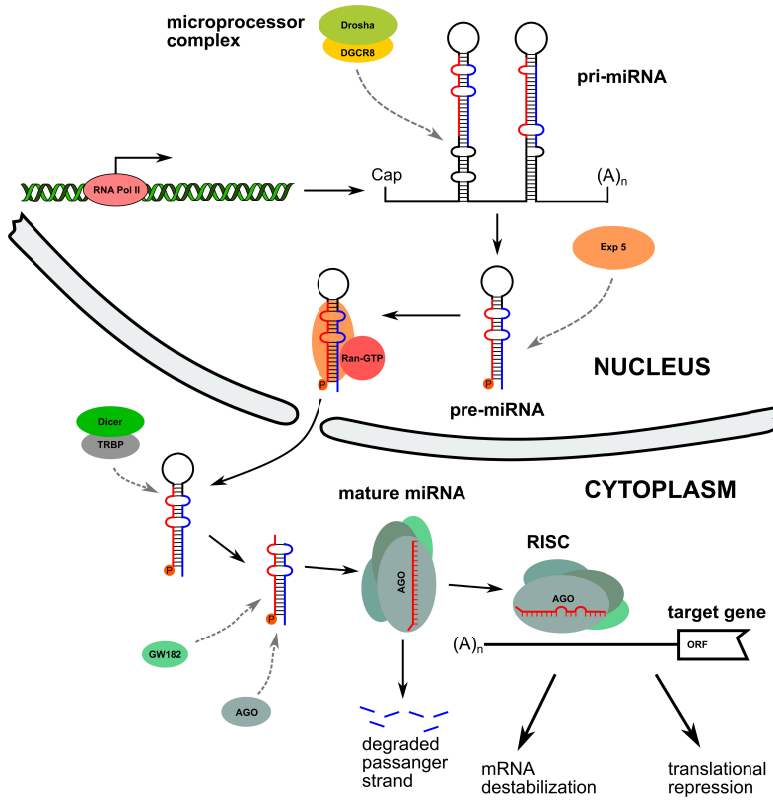


Figure 3. Canonical microRNA biogenesis pathway.

In RISC, miRNA duplex unwinding occurs. The biologically active "leading" or "guide" strand remains with the functional complex, whereas the inactive "passenger" or "star" strand is cleaved. In animals, the strand preferably chosen as the guide strand has a thermodynamically more unstable 5' end and uracil as the first nucleotide (Khvorova et al., 2003; Schwarz et al., 2003). In many cases, however, the passenger strand is also stable and functional (Okamura et al., 2008). Alternative arm selection or "arm-switching" events result in differential expression profiles of the miRNA arms in different tissues and developmental stages (Chiang et al., 2010), which can influence the corresponding targetome (Marco et al., 2012).

Whereas the overwhelming majority of miRNAs are preprocessed by the canonical pathway, some functional miRNAs have a slightly different biogenesis. Well-known examples include the mirtrons – miRNAs that are processed independently of the microprocessor complex (Berezikov et al., 2007; Okamura et al., 2007; Ruby et al., 2007). These miRNAs originate from introns of the host genes and are processed out of transcript by spliceosome. The resulting lariat is linearized by RNA lariat debranching enzyme (DBR1) and, if

needed, trimmed from the 3' or 5' end (3'- or 5'-tailed mirtrons) (Ruby et al., 2007). The rest of the biogenesis pathway is identical to the canonical type.

For at least one known miRNA, miR-451a, processing by Dicer is skipped. Instead, AGO2 performs cutting of the pre-miR-451a (Cheloufi et al., 2010; Cifuentes et al., 2010; Yang and Lai, 2010). Subsequently, the 3'-5' poly(A)-specific exonuclease PARN trims the remaining part of the pre-miRNA (Yoda et al., 2013), leaving a short 18-nt functional miRNA.

Another less-characterized noncanonical class of miRNAs are the intronic simtrons – small RNAs that are preprocessed independently of DGCR8, Dicer and spliceosome (Havens et al., 2012). Fraction of similarly sized small RNAs originate from noncoding RNAs, such as snoRNAs (Ender et al., 2008), tRNAs (Maute et al., 2013), endogenous small interfering RNAs (endo-siRNAs) (Babiarz et al., 2008; Tam et al., 2008; Watanabe et al., 2008), and small hairpin RNAs (shRNAs) (Babiarz et al., 2008), all of which are processed independently from Drosha/DGCR8. Although the origin and biogenesis of these small RNAs differ substantially from those of canonical miRNAs, they are able to form complexes with RISC and, in principle, regulate gene expression. For these reasons, some of them are classified together with miRNAs.

In addition to 5' and 3' miRNAs, a specific class of miRNA-offset RNAs (moRs) are frequently processed from the proximal regions of miRNA precursors (Langenberger et al., 2009; Shi et al., 2009; Zhou et al., 2012). However, the relative abundance of moRs is low compared to canonical mature forms (Zhou et al., 2012). As a result, their functions, action mechanisms, and relevance are not well understood. These exceptions illustrate that the rules in biological systems are often flexible, such that several independent ways can be used to address a common task.

Second-generation sequencing experiments have demonstrated that there is often some variability in the sequences of mature miRNA derived from the same pre-miRNA. Variants that differ from the miRBase reference sequence are termed isomiRs (Morin et al., 2008), and they are generated through several mechanisms. Processing steps by Drosha and Dicer are not completely precise and produce some portion of miRNAs with variable 3' or 5' ends (templated additions or trimmed ends in mature sequence) (Morin et al., 2008). MicroRNAs can be imprecisely trimmed by exonucleases like Nibbler, resulting in a population of variable miRNAs (Han et al., 2011). Nontemplate additions can be explained by the action of terminal nucleotidyl transferases (TNTases), which add specific nucleotides to the end of mature miRNA (Burroughs et al., 2010). RNA A-I editing by ADAR enzymes changes the internal sequence of some miRNAs (Kawahara et al., 2007). Finally, single-nucleotide polymorphisms (SNPs) can cause mature miRNA to differ from the reference sequence (Gong et al., 2012). Whereas changes in the 3' end of miRNA are expected to have a minor effect, differences in the 5' end can substantially influence the targetome of the miRNA.

1.1.3. Mode of action

According to current knowledge, miRNAs most commonly regulate gene expression in animals by binding to the 3' UTR of target mRNA (reviewed in; (Bartel, 2009)). This binding results in cleavage, translational arrest, or destabilization and subsequent decay of the mRNA. Cleavage of mRNA, which is mediated by the RNAse III class enzyme AGO2, only occurs when there is perfect complementarity between the mRNA and miRNA. Cleavage is a prevalent mechanism in plants but rarely reported in animals (Yekta, 2004).

Translational arrest was long thought to be the main mechanism of miRNA-mediated regulation in animals. Several models of action have been proposed and supported by experimental data, but the debate about the relevance of each of those is still ongoing. For example, miR-RISC has been shown to recruit the eIF6 antiassociation factor, which, in turn, inhibits association of the 60S subunit and subsequent formation of functional ribosome (Chendrimada et al., 2007). Similarly, several studies have demonstrated that miR-RISC inhibits recognition of the m7-G cap-structure by eIF4E (Humphreys et al., 2005; Pillai et al., 2005). Additionally, miRNAs have been proposed to have a postinitiation inhibitory effect on translational elongation, causing ribosomal drop-off and cotranslational protein degradation (Petersen et al., 2006).

Substantial evidence supports the idea that regulation by miR-RISC has a destabilizing effect on mRNA (Bagga et al., 2005), through the combined effect of poly(A) tail removal (Giraldez et al., 2006), mRNA decapping (Behm-Ansmant et al., 2006; Rehwinkel et al., 2005), and subsequent 5'-3' degradation. The miR-RISC complexes with bound targets are sometimes sequestered to specific foci, P-bodies, which are enriched by enzymes functioning in mRNA degradation (Liu et al., 2005a, 2005b). mRNA destabilization and subsequent degradation is the main contributor to the miRNA-mediated reduction of protein level (Eichhorn et al., 2014; Guo et al., 2010) and is usually coupled with the more modest effect of translational arrest.

Although miRNAs predominantly function through binding to the 3' UTR, in silico predictions and experimental data show that miRNAs can also bind to the 5' UTR (Lytle et al., 2007; Miranda et al., 2006) and coding DNA sequence (CDSs) (Tay et al., 2008). However, targeting in those regions does not confer as large of an effect on target expression as does targeting the 3' UTR, but rather complements the latter (Fang and Rajewsky, 2011; Grimson et al., 2007).

In some cases, miRNAs may regulate gene expression in ways which are quite different from overall logic. For example, in rare cases they can upregulate (rather than downregulate) the expression of target genes (Vasudevan et al., 2007). There are also reports that some miRNAs exert their regulatory effect on the DNA level, by activating (Place et al., 2008) or silencing transcription (Kim et al., 2008).

In contrast to plants, miRNA-mediated regulation in animals usually involves imperfect complementarity between miRNA and mRNA. The main contributor of animal miRNA action is the amount of complementarity between

the 5’ end of the miRNA and miRNA target site (Lewis et al., 2003). This critical miRNA recognition element (MRE) is designated as the “seed” region and comprises nucleotides 2 to 7 from the 5’ end of miRNA (Brennecke et al., 2005; Lewis et al., 2003, 2005). Approximately 80% of all experimentally supported miRNA target sites have some degree of complementarity with the MRE (Grosswendt et al., 2014). Canonical miRNA target sites are classified based on the length of perfect pairing (Fig. 4). Whereas minimal MREs have perfect pairing only with the miRNA seed region (6mer), more efficient canonical MREs have adenine adjacent to the first position of miRNA (7mer-A1), additional pairing in position 8 of the miRNA (7mer-m8), or both features (8mer) (Lewis et al., 2005).

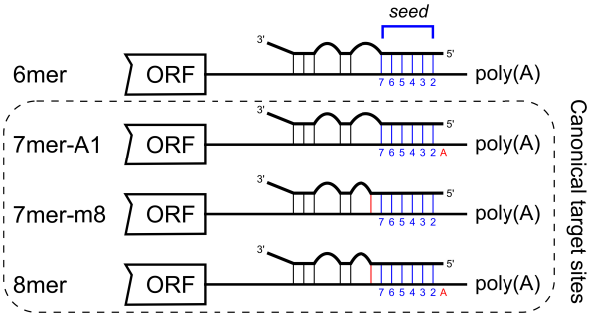


Figure 4. Schematic representations of canonical miRNA target sites in order of increasing effectiveness.

Several studies have described functional target sites that do not adhere to the requirement of perfect pairing with the seed region. Examples of these noncanonical MREs involve “centered” target sites with complementarity in the central part of the miRNA. Centered MREs typically affect target on the protein level (Martin et al., 2014). High-throughput experiments have shown that some MREs with wobble, mismatch, or bulge in the seed region overlap with RISC binding sites (Chi et al., 2012; Hafner et al., 2010; Helwak et al., 2013). For example, G-bulge sites, which have a bulging G nucleotide in the seed region, could comprise more than 15% of putative MREs identified from the HITS-CLIP dataset of mouse brain (Chi et al., 2012). Mismatches in the seed region can be compensated (to some extent) by pairing in nucleotides 13–16. However, these so-called 3’-compensatory sites (Friedman et al., 2009) represent a small proportion of all predicted binding sites (Bartel, 2009). Although noncanonical target sites seem to be widespread (Helwak et al., 2013), their effects on the mRNA and protein levels are weaker than those of canonical MREs (Khorshid et al., 2013; Wang, 2014).

According to estimates from in silico target prediction algorithms, more than 60% of all human protein-coding genes have been under positive selection to

maintain pairing with miRNAs (Friedman et al., 2009) and are, therefore, likely targets. One animal miRNA may regulate hundreds of targets, and one target may be regulated by several miRNAs. Hence, the miRNA regulatory networks are very complex. Furthermore, miRNA action depends on the spatiotemporal expression profile and cellular context (e.g., targetome, regulating transcription factors, and competing effects of other miRNAs and RNA-binding proteins [RBPs]). As a result, it can be difficult to untangle the functions of one particular miRNA. A single MRE typically has a modest effect on protein level (Baek et al., 2008), but the cooperation of several binding sites can have an enhanced impact (Doench and Sharp, 2004; Grimson et al., 2007).

In addition to known cellular pathways and processes under the regulation of specific sets of miRNAs (e.g., hypoxia pathway under regulation of miR-210-3p, as reviewed in; (Chan et al., 2012; Huang et al., 2010)), many miRNAs have more general function of ensuring the robustness of transcriptional programs. MicroRNAs frequently participate in incoherent feedforward and negative feedback loops, which can buffer the noise arising from stochastic events, such as leaky transcription (reviewed in; (Ebert and Sharp, 2012)). The transcriptome of many tissues tends to be depleted of the MREs of miRNAs expressed in the same tissue. This "target avoidance" phenomenon is necessary for maintaining tissue specificity and developmental timing (Farh et al., 2005; Stark et al., 2005).

Taken together, these previous studies suggest that miRNAs participate in tight spatiotemporal control over the expression of mRNAs and proteins. This control, in turn, is a prerequisite for correct functioning of cellular systems in complex organisms.

1.2. MicroRNA profiling and analysis strategies

Several strategies for identifying the biological functions of miRNAs have been developed. This section discusses the main methodological approaches used in miRNA research.

1.2.1. MicroRNA expression profiling

During the past 15 years, three main methods have been routinely used for profiling the expressions of miRNA and mRNA. These methods have been modified to accommodate the technical problems arising from the small size of mature miRNAs.

1.2.1.1. Expression arrays

Expression array technologies are based on hybridization between fluorescently labelled target sequences and complementary detection probes attached to a

glass slide or bead. After removal of the unhybridized material, the quantity of hybridized sequences is determined by the strength of the light signal in each set of identical detection probes. Signals are detected and quantified by a specialized scanner, resulting in numerical values. Subsequent preprocessing (e.g., quality control, filtering, transformation, normalization) and statistical analysis methods vary depending on the specific array platform, dataset, study design, bioinformatics software package, and biological question.

The short length of miRNAs limits options for optimizing the melting temperature (T_m) over hundreds of detection probes. However, this consideration is crucial for avoiding biases in binding efficiencies. Several strategies have been used to address this issue. For example, locked nucleic acids (LNAs) can be incorporated into detection probes to achieve more stable T_m values over the array (Castoldi et al., 2006). Another strategy is to optimize the length of each detection probe (Baskerville and Bartel, 2005).

The main advantage of expression arrays is their relatively low cost for global miRNome profiling compared to other methods. Disadvantages include their limited dynamic range and ability to detect only the specific set of miRNAs used in array design (usually based on current miRBase version). Thus, novel miRNAs and isomiRs remain undetected. Due to the recent decrease in cost and substantial benefits of massively parallel sequencing, microarray-based methods are expected to become outdated in coming years.

1.2.1.2. Small RNA sequencing

The first step of small RNA sequencing methods is preparation of a cDNA library from a small RNA sample, followed by parallel sequencing of fragments in the library. Standard bioinformatics preprocessing steps include removal of sequencing adapters, quality control, and read filtering. Preprocessed reads are aligned to the reference genome, and to the mature and precursor miRNA sequences. Expression of each miRNA is determined by counting the number of reads that map to the region of mature miRNA. This count is used in subsequent preprocessing and analysis steps. As these technologies determine the discrete counts of reads, rather than the intensities of light signals, most of the widely used normalization and/or statistical analysis methods differ substantially from methods used in expression array analyses (compared in; (Rapaport et al., 2013; Soneson and Delorenzi, 2013)).

The main advantage of small RNA sequencing is that, in addition to miRNA expression quantification, acquired data can be used to detect isomiRs and identify novel miRNAs (Morin et al., 2008). Important disadvantages are the requirement for a larger amount of RNA input, a complicated library preparation procedure, and a considerably higher cost compared to array-based technologies.

1.2.1.3. Quantitative reverse transcription – polymerase chain reaction (qRT-PCR)

The qRT-PCR method is based on reverse transcription (RT) of miRNAs to synthesize cDNA, followed by quantitative polymerase chain reaction (qPCR) to quantify the accumulated product by measuring the light emitted in each round of amplification. To accommodate qRT-PCR to the short length of target sequences, specific stem-loop primers can be used in the RT phase (Chen et al., 2005). An alternative strategy involves polyadenylation of the miRNA before the RT step and subsequent use of universal primers. The qRT-PCR approach has been used primarily for quantifying specific preselected miRNAs and validating results of expression arrays or small RNA sequencing experiments. However, panels of many miRNA primers are available in microfluidic plate format for global profiling of many miRNAs. The T_m optimization strategies for qRT-PCR panels are similar to those used for miRNA expression arrays.

The major advantage of qRT-PCR is that it offers the highest detection sensitivity among available methods, making it the method of choice when the quantity of input material is low. Disadvantages include the possibility of obtaining quantifications only for a set of preselected miRNAs and a higher price compared to expression arrays.

1.2.2. Target prediction algorithms

Substantial effort has been made towards developing bioinformatic methods to predict targets of miRNAs of interest. Numerous target prediction methods have been developed for animals and plants. However, because target regulation follows different rules in plants versus animals (i.e., in plants, target-miRNA complementarity must be complete or near-complete, and there are no preferential target site locations in 3' UTRs), I focus only on animal-specific approaches in this thesis. Target prediction for animal miRNA targets is also computationally more complex, because only partial complementarity between the target and miRNA is necessary for regulation.

Perfect or near-perfect complementarity between the miRNA seed region and target is a major contributor to miRNA functionality (Lewis et al., 2003, 2005). Consequently, the easiest and most naïve way to predict putative miRNA targets would be to search for perfectly complementary sites between the minimal 6mer seed region of miRNA and the target transcript. However, as this approach yields many false positives, additional information is necessary to enhance the usefulness of in silico target prediction.

Some frequently used additional criteria for target prediction are as follows:

1. Non-seed complementarity: Larger complementarity outside the seed region, in the 3' end of miRNA, enhances miRNA effectiveness (Grimson et al., 2007).
2. Conservation: Similarly to other genomic elements, MREs that are conserved in several species are more likely to be functional. However, using

this threshold as a strict filter misses the large number of nonconserved but functional target sites (Farh et al., 2005).

3. Free energy: Higher stability of the miRNA-mRNA duplex is associated with a stronger effect of MRE.
4. Target site accessibility: For miRNA-mediated gene regulation to be functional, the secondary structure of the target must permit binding of the miRNA and accompanying protein complex.
5. Local AU percentage: Greater AU nucleotide content near the functional binding site has been correlated with greater strength of the MRE (Grimson et al., 2007), probably due to the more favorable secondary structure of mRNA and, therefore, better accessibility of the binding site.
6. Number of target sites: Presence of multiple binding sites in one 3' UTR enhances the functionality of a given miRNA (Grimson et al., 2007).
7. Target site location and distribution: Empirical evidence indicates that MREs positioned near the 5' or 3' end of the 3' UTR are functionally stronger than those positioned in the middle (Grimson et al., 2007). Additionally, pairs of closely positioned target sites have a synergistic effect on gene regulation (Grimson et al., 2007).
8. Pairing stability in miRNA seed region: MicroRNAs having globally weaker seed pairing stability with their targetome have a smaller proficiency of gene regulation (Garcia et al., 2011).
9. Target abundance: MicroRNAs targeting a larger number of genes have smaller effectiveness of gene regulation (Garcia et al., 2011).

Earlier prediction algorithms used a rule-based approach to identify putative miRNA binding sites. These algorithms were based on features defined from information gathered through low-throughput experiments. Some newer implementations use more complex approaches and experimentally validated data to train the algorithms. The training of the algorithm can involve several machine-learning strategies, such as support vector machines (Wang and El Naqa, 2008) or artificial neural networks (Chandra et al., 2010). However, some of the datasets used for defining rules or training algorithms may not be representative, due to the small number of confidently validated functional and nonfunctional miRNA-target interactions (Fan and Kurgan, 2015). Most algorithms have been implemented for predicting target sites from the 3' UTRs of genes. Nevertheless, some of these algorithms can also be used for target prediction from the coding regions of genes (by using the full transcript as input), whereas others implement this feature exclusively (e.g., DIANA microT-CDS). In the following subsection, I will discuss the principles of some of the most widely used target prediction algorithms.

1.2.2.1. Overview of the algorithms

1.2.2.1.1. Method of Stark et al.

The first study to utilize the in silico method for miRNA target prediction sought to identify targets of the *Drosophilia melanogaster* miRNome (Stark et al., 2003). Authors screened 3' UTRs for complementarity between the mRNA and first eight nucleotides from the 5' end of miRNA. Differences between the observed free folding energy (ΔG) and binding energies of random sequences were measured by Z-scores, with a strict criterion of $Z \geq 3$ being used as a threshold for conservative target prediction.

1.2.2.1.2. TargetScan, context+ score, and probability of preferentially conserved targeting (P_{CT})

One of the first target prediction algorithms to be developed (Lewis et al., 2003, 2005), TargetScan remains one of the most popular strategies for miRNA target prediction. This simple algorithm performed relatively well in independent test sets (Fan and Kurgan, 2015). The method searches for matches between the miRNA seed region (8mer, 7mer-A1, 7mer-m8) and the 3' UTR of the target gene. Conservation of the target site across species is used as an additional filter for target site identification (although prediction without the conservation filter is possible). The original algorithm (Lewis et al., 2003) considered free binding energy constraints, pairing efficiency outside the seed region, and multiple binding sites. However, these criteria were dropped in the current implementation, denoted as TargetScanS (Lewis et al., 2005). The resulting precomputed target database has been renewed several times (v7.0 is the latest version, at the time of writing). The algorithm also has been used to predict targets from CDSs (Lewis et al., 2005).

In conjunction with the TargetScan algorithm, the context+ score has been widely used for the prioritization of target sites (Garcia et al., 2011; Grimson et al., 2007). This score combines contributions from six contextual miRNA binding site features that correlate with targeting efficiency and were identified in gene downregulation data from miRNA transfection experiments (Grimson et al., 2007). These features are the site type (7mer-A1, 7mer-m8, 8mer), pairing in the 3' region of miRNA, local AU quantity, binding site location, global miRNA binding site abundance, and miRNA seed pairing stability. Smaller context+ score values indicate more favorable binding and a higher possibility that the miRNA-target pair is functional and biologically relevant. The recent update of the score (context++ score) improves the predictor even further, using 14 contextual features (Agarwal et al., 2015). Finally, the P_{CT} method (Friedman et al., 2009) can be used to identify MREs that are selectively maintained in the evolution due to miRNA targeting and, therefore, potentially more relevant.

1.2.2.1.3. DIANA-microT algorithms

The first iteration of the DIANA-microT algorithm was developed at the same time as TargetScan. To develop this algorithm, the authors carried out experiments with putative miRNA-mRNA targets to define features that influence miRNA repression strength (Kiriakidou, 2004). The initial algorithm used dynamic programming to identify the mRNA-miRNA duplex with a minimal binding energy threshold. Additional requirements included perfect or near-perfect complementarity in the 5' end of the miRNA-mRNA duplex (up to two wobble pairs allowed), at least 5-nt-long canonical or wobble pairing in the 3' end of the duplex, and the presence of a central bulge or loop.

A major modification of the method, DIANA-microT v3.0 (Maragkakis et al., 2009a, 2009b) uses dynamic programming to identify the best-scoring alignment of 6- to 9-nt binding in the 5' end of the miRNA and 3' UTR of the target. For perfectly complementary regions that are smaller than 7 nt (including 7–9 nt regions with wobble pairs), additional constraints for free binding energy are used. This additional filter is achieved by using shuffled "mock" miRNAs for defining the background and identifying target sites that have significantly lower free binding energy than the background. Species conservation information is used to score each MRE (compared to MREs of the mock miRNAs). Individual scores per gene are used to calculate the miRNA target gene score (miTG).

DIANA-microT v4 (Maragkakis et al., 2011) enhanced the previous algorithm by using available pulsed stable isotope labeling with amino acids in cell culture (pSILAC) data (Selbach et al., 2008), instead of shuffled mock miRNAs, for defining the nonfunctional background. The latest version, DIANA-microT-CDS (Paraskevopoulou et al., 2013; Reczko et al., 2012), is one of the few algorithms to address specifically miRNA target site identification from the CDS, in addition to 3' UTR. This algorithm uses dynamic programming for putative MRE identification. Target site conservation, target site accessibility, local AU content, distance from the 3' UTR end, target site distance from other target sites, free binding energy, and miRNA binding pattern are used as additional features for MRE scoring. Feature selection, target site scoring, and binding site prioritization were carried out by using available PAR-CLiP data (Hafner et al., 2010).

1.2.2.1.4. miRanda and miRSVR

The miRanda algorithm (Enright et al., 2003; John et al., 2004) aligns the miRNA sequence against the 3' UTR and uses alignment quality scores to evaluate binding strength. Subsequently, free binding energy is calculated and used to filter the results. The algorithm can be used with or without the requirement of strict complementarity in the 7-nt seed region. Additional conservation filters are added for the precomputed results (http://www.microrna.org/).

Together with miRanda algorithm, the miRSVR score can be used as an additional measure of binding site efficiency (Betel et al., 2010). This measure was constructed by using the machine-learning approach of support vector

regression, in which the model was trained on gene expression data from miRNA transfection experiments (Grimson et al., 2007). Several binding site and sequence context features were used for training, as follows:

MRE features
1. Complementarity with the seed region
2. Complementarity in the 3' part of the miRNA

Local context features
3. AU nucleotide quantity in the MRE-flanking region
4. Secondary structure accessibility

Global context features
5. Length of the 3' UTR
6. MRE distance from the UTR ends
7. Conservation level of the MRE region

A lower (negative) miRSVR score indicates a higher probability that the binding site is functional. Unlike the context+ score, miRSVR is also useful for prioritizing noncanonical binding sites.

1.2.2.1.5. PicTar

PicTar (Krek et al., 2005) uses the 7-nt seed requirement, conservation, and free energy filters for identifying a set of highly probable binding sites, called anchors. If the 3' UTR has enough anchors, then a hidden Markov model is used to calculate the maximum likelihood that the corresponding 3' UTR will be targeted by some combination of input miRNAs. Therefore, PicTar specifically addresses the combinatorial effect of miRNAs.

1.2.2.1.6. Probability of interaction by target accessibility (PITA)

The PITA (Kertesz et al., 2007) algorithm assesses miRNA-mRNA complementarity and the accessibility of putative miRNA target sites for RISC. The method searches for 6- to 8-nt complementary seed regions from the target 3' UTR. Dynamic programming is used to calculate the minimum free energy of the putative miRNA-mRNA duplex (ΔG_{duplex}) and the energy required to make the binding site accessible (ΔG_{open}). The difference between these two energies ($\Delta\Delta G$) is used to prioritize individual target sites. Additionally, the effect of multiple binding sites of one miRNA is calculated by the formula:

$$T = log \sum_{i=1}^{n} e^{s_i}$$

where s_i is the $\Delta\Delta G$ of an individual binding site; and n is the number of binding sites for the corresponding miRNA.

1.2.2.2. Performance of target prediction algorithms

In silico target prediction algorithms are useful for finding putative miRNA-target interactions and are usually the first step when interpreting the functions of miRNA(s) of interest. However, these methods also have limitations, such as a limited ability to identify truly functional miRNA-target pairs and some proportion of false positives. Attempts have been made to assess the performances of target prediction algorithms (Maragkakis et al., 2009c; Rajewsky, 2006; Sethupathy et al., 2006). For accurate analysis of performance, independent (i.e., not used for algorithm training), reasonably large, and confidently identified sets of true and false miRNA-target associations are needed. Experimentally validated miRNA-target pairs or MREs supported by AGO-CLIP have been routinely used as true associations.

Algorithm performance can be assessed by several measures. Sensitivity indicates the proportion of correctly predicted miRNA-target interactions among all true associations, whereas specificity indicates the proportion of false miRNA-target interactions that are correctly identified as false. Precision indicates the proportion of identified miRNA targets that are true targets. A good prediction algorithm should combine high sensitivity, specificity, and precision. The ability of prediction algorithms to pinpoint true target associations is often evaluated by the receiver operator characteristic (ROC) curve, wherein the x-axis denotes 1-specificity and the y-axis denotes sensitivity. In the case of a continuous predictor indicating the probability of a real interaction (e.g., context+ or miRSVR score), the sensitivity and specificity are calculated for each predictor value and plotted as a curve. A high value for the area under the ROC curve (AUC) indicates good algorithm performance; an AUC of 0.5 indicates poor performance and random classification.

Fan and Kurgan carried out the most comprehensive comparative analysis of prediction algorithms to date (Fan and Kurgan, 2015). After reviewing 38 algorithms designed for miRNA target prediction in animals, the authors comprehensively evaluated the performances of seven algorithms: TargetScan, DIANA-microT-CDS, miRanda, PicTar, miRmap, miRtarget2, and EIMMo. Confidently validated miR-target interactions (by qRT-PCR, luciferase assay, or Western blot) from miRTarbase (Hsu et al., 2014) were used to construct corresponding sets of functional and nonfunctional pairs. Unlike similar studies, the authors exclusively used interactions that were validated after the newest tested prediction algorithm was released, to rule out potential overlap between the test set used for evaluation and training sets used for prediction algorithm optimization.

The authors reported significantly better performances for algorithms that made predictions at the target gene level (AUC 0.59–0.75), compared to those making predictions at the binding site level (AUC 0.52–0.67). Among the tested algorithms, TargetScan and DIANA-microT-CDS had good overall performances in most aspects. PicTar (AUC 0.59 and 0.54 for gene and binding site level)

had relatively low sensitivity (0.34 and 0.27) but high specificity (0.83 and 0.80), meaning that the predicted target sites were more likely to be true MREs.

One widely used strategy to identify more credible set of miRNA targets is to intersect the results of several target prediction algorithms. This approach will reduce the sensitivity of the prediction, while making the results more conservative by increasing the specificity. Additionally, overlapping the in silico predicted MREs with experimentally determined RISC binding sites and prioritizing targets based on miRNA binding efficiency measures (e.g context+ and miRSVR scores) can increase the specificity of target prediction algorithms.

In conclusion, algorithms utilizing different and often quite sophisticated methods for target prediction can serve as tools for narrowing down the potential targetome and obtaining an overall view of the functions of the miRNA(s) of interest. The most interesting interactions having potential impact on cellular networks, traits, or pathogenic changes should be validated experimentally before they are declared to be drug targets or biomarkers.

1.2.3. Experimental techniques for miRNA target identification

Various experimental techniques are available for identifying miRNA targets and complementing the results of silico target predictions. Experimentally supported mRNA-miRNA interactions have been collected in databases, such as TarBase (Vlachos et al., 2015), miRTarbase (Chou et al., 2016), and miRecords (Xiao et al., 2009). Methods can be classified based on scope (low- vs. high-throughput) or level of biological information (effect on mRNA vs. protein level).

1.2.3.1. Reporter gene assays

Often considered to be the gold standard for miRNA target validation, reporter gene assays are used to determine whether an interaction exists between the miRNA and target sequence (Lewis et al., 2003). The target region is cloned into a reporter vector, downstream of the open reading frame (ORF) of the gene encoding a fluorescent protein (e.g., luciferase). Cotransfection with an miRNA mimic targeting the region will result in reduced emission of light by the fluorescent reporter protein, indicating a functional interaction between the miRNA and tested target. Similarly, it is possible to downregulate an endogenously expressed miRNA by using chemically modified anti-miR oligonucleotides (AMOs), and then test the effect on target sequence regulation. However, this is still model system with its limitations – many factors may influence miRNA binding under physiological conditions.

1.2.3.2. Gene expression profiling after modulation of the miRNA level

A commonly used strategy to determine the effect of an miRNA on the transcriptome level is to use gene expression profiling methods after transfection with an miRNA mimic or after downregulation of the miRNA by AMOs. This strategy can be implemented in a high-throughput (e.g., utilizing microarrays or massively parallel sequencing techniques) or low-throughput manner (e.g., utilizing qRT-PCR or Northern blotting to determine expression changes of a specific target gene). However, this strategy cannot distinguish the direct effect of miRNA binding from secondary effects, such as those that emerge from the interplay between genes in regulatory networks and loops.

1.2.3.3. Determining protein abundance after modulation of the miRNA level

To determine the effects of miRNA targeting on the protein level, protein abundance can be measured after transfection with an miRNA mimic or AMO. Low-throughput strategies involve Western blotting, enzyme-linked immunosorbent assay (ELISA), fluorescent-activated cell sorting (FACS), and other standard proteomics methods for detecting the presence and abundance of the target protein. In high-throughput settings, pulsed stable isotope labeling in cell cultures (pSILAC) have been used (Selbach et al., 2008). This method uses the marking of newly synthesized proteins with isotope-labelled amino acids prior to mass-spectrometry, enabling to investigate the effect of miRNA on proteome.

1.2.3.4. Pull-down of biotin-marked miRNA

To identify the interaction partners of specific miRNAs, one strategy is to label the 3' end of synthetic miRNA with biotin and, after transfection, to use streptavidin-coated magnetic beads to purify the miRNA together with targeted transcripts (Martin et al., 2014; Orom et al., 2008). Resulting targets are analyzed by expression arrays or massively parallel sequencing.

1.2.3.5. Immunoprecipitation of RISC components

RISC proteins, such as AGO2, can be immunoprecipitated together with bound miRNAs and fragments of target mRNAs. Subsequent massively parallel sequencing methods can identify the locations of RISC binding and the array of miRNAs that are bound to RISC. Some methods, such as high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLiP) (Chi et al., 2009) and photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLiP) (Hafner et al., 2010), use ultraviolet (UV) treatment to crosslink RISC proteins with RNA. The limitation of these

methods is that they do not provide explicit information about the miRNA-mRNA interaction in the region of RISC binding and this is usually inferred from in silico target prediction. The crosslinking, ligation, and sequencing of hybrids (CLASH) method (Helwak et al., 2013) ligates miRNA and mRNA before sequencing, enabling miRNA targets to be mapped regardless of in silico target prediction algorithms.

1.3. MicroRNAs in cancer

During the last 15 years, miRNA-mediated regulation has been shown to be relevant for many biological processes, with impairments being associated with several pathologies, such as heart (van Rooij et al., 2006; Wahlquist et al., 2014), neurodegenerative (Miñones-Moyano et al., 2011), and autoimmune diseases (Junker et al., 2009). However, the main focus of miRNA research has been on their involvement in tumorigenic processes.

Soon after the realization that gene regulation via miRNA targeting is a widespread and evolutionally conserved mechanism, it was reported that the region encoding two miRNAs (cluster of miR-15 and miR-16 in 13q14.3) is frequently deleted in the chronic lymphocytic leukemia (CLL) (Calin et al., 2002). After this observation, more comprehensive bioinformatic screening was conducted for all miRNAs known by that time (Calin et al., 2004). Out of the 186 miRNA genes known by then, 35 mapped into fragile sites frequently mutated in cancers, indicating significant overrepresentation (P<0.0001). Additionally, about half of the miRNA genes were mapped to the same chromosome band as the established fragile site. A similar association was found for mouse cancer susceptibility loci (Sevignani et al., 2007). However, these studies analyzed only the subset of miRNAs known to date and later research suggested that preferential location near fragile sites is not specific to miRNAs, but also applies to protein-coding genes (Laganà et al., 2010).

Subsequently, large-scale miRNA profiling studies confirmed that the expression of many miRNAs is changed in several cancer types (Iorio et al., 2005; Lu et al., 2005; Volinia et al., 2006; Yanaihara et al., 2006). Their expression profiles were found to distinguish cancerous from noncancerous tissues and to classify poorly differentiated samples more accurately than mRNA profiles (Lu et al., 2005). Similarly to protein-coding genes, the expression of miRNA genes in cancer can be dysregulated by genomic deletions or duplications, changes in the methylation patterns of promoters, and regulation by cancer-associated transcription factors. Cancer-associated mutations in individual components of miRNA processing machinery are also frequently reported. Such mutations can cause global changes in miRNA levels and subsequent dysregulation of the whole transcriptome (Melo et al., 2010). Incidentally, global downregulation of the miRNome has been reported in cancerous tissues (Lu et al., 2005). Given that miRNAs “fine-tune” the gene expression, this downregulation may have a destabilizing effect on the operation

of the whole cellular machinery. Additionally, variants in miRNAs (Rawlings-Goss et al., 2014) or their binding sites (Chin et al., 2008; Nicoloso et al., 2010) can contribute to the formation of tumorigenic traits, by mechanisms that are discussed in detail in the next chapter.

1.3.1. MicroRNAs as oncogenes or tumor suppressors

MicroRNAs may act as oncogenes (oncomiRs) or tumor suppressors by regulating corresponding cancer-associated genes or cellular pathways. The first functional evidence for the oncogenic function of miRNAs was for the six members of the miR-17-92 cluster (mir-17, -18a, -19a, -19b-1, -20a, and -92a-1), named oncomiR-1 (He et al., 2005b). Members of this cluster were up-regulated in B-cell lymphoma (BCL) samples and contributed to tumor development in a BCL mouse model. Numerous established tumor-suppressor genes, including *PTEN* (Xiao et al., 2008) and *E2F1* (O'Donnell et al., 2005), are experimentally supported targets of this cluster.

One of the most well-known examples of tumor-suppressor miRNAs is the let-7 family, members of which are downregulated in several cancer types (Ali et al., 2012; Dahiya et al., 2008; Takamizawa, 2004) and correlated with lung cancer survival (Takamizawa, 2004; Yanaihara et al., 2006). Among the validated targets of these miRNAs are established oncogenes from the *RAS* family (Johnson et al., 2005), *MYC* (Sampson et al., 2007) and *HMGA2* (Lee and Dutta, 2007; Mayr et al., 2007; Shell et al., 2007).

In therapeutic applications, oncomiRs can be downregulated by synthetic antagomiRs, chemically modified oligonucleotides that bind targeted miRNA and deactivate its biological function (Krützfeldt et al., 2005). Similarly, miRNA "sponges", synthetic RNA constructs with several MREs, can be used to decrease expression of targeted miRNAs (Ebert et al., 2007). Circular RNAs (circRNAs) are currently in the limelight of research, due to their function as natural "sponges" and their emerging biological relevance (Guo et al., 2014; Memczak et al., 2013; Salzman et al., 2012). These RNAs are more stable than linear sponges and can act as models for more effective therapy.

Mimics of tumor-suppressor miRNAs can be delivered to the location of action by viral vectors (Kota et al., 2009) or nonviral carriers, like atelocollagen (Tazawa et al., 2007) or liposome-polycation-hylaruronic acid nanoparticles (Chen et al., 2010). At the time of writing, first miRNA-based cancer drug MRMX4, developed by Mirna Therapeutics (Austin, Texas), is under Phase I clinical trials for use in several cancer types. MRMX4 is a mimic of a well-established tumor suppressor, miR-34a-5p, which is encapsulated into liposomal nanoparticle formulation (http://www.mirnarx.com/pipeline/mirna-MRX34.html).

1.3.2. MicroRNAs as cancer biomarkers

The ability to differentiate tissue samples based on cancer status (Volinia et al., 2006), type (Lu et al., 2005), subtype (Landi et al., 2010; Zhang et al., 2012b), or stage (Mascaux et al., 2009; Olson et al., 2009) underscores the potential utility of miRNA profiles as diagnostic biomarkers. Physicians are better equipped to make treatment decisions when they have a precise diagnosis. For example, whereas bevacumizab (Avastin) is effective for the treatment of lung adenocarcinoma, its introduction was associated with life-threatening hemorrhage in the case of squamous cell lung cancer (Johnson et al., 2004). Moreover, it is demonstrated that the miRNA profiles of metastases reflect the profiles of primary tumors (Rosenfeld et al., 2008) making it possible to "track down" the location and type of unknown primary tumor. As the miRNA profile may reflect the properties of a specific cancer, it may also correlate with the efficiency of a specific anticancer treatment. For instance, miR-520g-3p has been reported to mediate resistance to 5-fluorouracil and oxaliplatin therapy in colorectal cancer (Zhang et al., 2015a). Such information may help physicians to target treatments better, such as by choosing a more suitable drug or targeting the corresponding miRNA(s) with antagomiRs, in addition to standard treatment.

Expression profiles of miRNAs have been shown to correlate with disease progression or postoperative survival in several cancer types, suggesting their potential as prognostic biomarkers. The first such association was demonstrated in CLL, where the expression profiles of nine miRNAs correlated with the time to development of symptoms (Calin et al., 2005). Another landmark study in non-small cell lung cancer (NSCLC) found an association between reduced postoperative survival and high miR-155-5p expression (Yanaihara et al., 2006). More recent study demonstrated that the expression of individual miRNAs was not robust prognostic biomarker in breast cancer. However, the combined signature of 17 known and 24 novel miRNAs was more precise, dividing breast cancer samples into the different prognostic groups (Dvinge et al., 2013).

The most exciting property of miRNAs is that they are stable in several body fluids, including blood (Mitchell et al., 2008), sputum (Xie et al., 2010), and urine (Hanke et al., 2010; Weber et al., 2010). In blood, miRNAs are protected from RNAse digestion by being bound with AGO proteins (Arroyo et al., 2011), encapsulated in exosomes or microvesicles (Valadi et al., 2007), bound by high-density lipoprotein particles (Vickers et al., 2011), or incorporated in apoptotic bodies (Zernecke et al., 2009). Cell-free miRNAs can be the byproducts of cell death, or they may be secreted actively and selectively (Guduric-Fuchs et al., 2012; Pigati et al., 2010), representing an intriguing form of cell-cell communication. Either way, they can serve as valuable noninvasive biomarkers, as the expression profiles of miRNAs in plasma and serum are altered in several cancer types.

The first study demonstrating the potential utility of blood miRNAs as cancer biomarkers showed that the plasma expression of miR-21-5p was

associated with relapse-free survival in BCL (Lawrie et al., 2008). Since then, blood profiles of different miRNAs have been associated with cancer diagnosis and prognosis in various studies in lung (Boeri et al., 2011; Chen et al., 2008), breast (Roth et al., 2010), and prostate cancers (Mitchell et al., 2008). Effective noninvasive biomarkers would enable physicians to diagnose cancer (sub)types, and to predict disease and treatment outcomes in an efficient and patient-friendly way. Several commercial miRNA-based diagnostic assays are already available. For example, Rosetta Genomics™ offers miRNA-based tests for better diagnosis of lung, thyroid, and kidney cancers, as well as for identifying the primary origin of tumors (Meiri et al., 2012).

1.3.3. MicroRNAs in lung cancer

In 2012, lung cancer was the most commonly diagnosed cancer and the leading cause of cancer-related deaths worldwide (Ferlay et al., 2015). Most lung cancers are classified into two histologically different groups: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). The latter group is classified into three subtypes: squamous cell carcinoma (SCC), adenocarcinoma (AD), and large cell carcinoma (LCC). SCLC constitutes about 10–15% of all lung cancer cases (Houston et al., 2014), has a very poor prognosis, and is generally inoperable. SCC comprises about 30% of lung cancer cases (Houston et al., 2014), originates from squamous cells of the inner lining of the lungs, and is strongly associated with smoking. AD constitutes about 40% of lung cancer cases (Houston et al., 2014) and originates from the glandular epithelium of the lung. AD has a slightly better prognosis and constitutes a higher percentage of never-smoker cases. Lung cancer types investigated in this thesis belong mainly to the SCC and AD subtypes of NSCLC.

By far, the greatest contributor to lung cancer risk is smoking history (~90% of all lung cancer patients have a history of smoking). Additional established risk factors involve exposure to second-hand smoke, asbestos, and radon (reviewed in; (Molina et al., 2008)). There is also a genetic component to lung cancer risk. Specifically, region 15q25.1 has been associated with lung cancer risk in genome-wide association studies (GWASs) (Thorgeirsson et al., 2008), possibly through modulation of smoking behavior.

Although the overall 5-year survival rate of NSCLC is poor (~20%; SEER Cancer Statistics Review 1975-2012, period of 2005–2011), patients with early IA stage show a survival rate up to ~80% in some populations (Goya et al., 2005). This fact indicates the need for better diagnostic and prognostic biomarkers to discover and diagnose early stages of disease, as well as more efficient drugs for the postoperative cure of patients with later-stage tumors.

The first association between miRNA and lung cancer was found in 2004 (Takamizawa, 2004). Five primary sequences of let-7 family miRNAs were shown to have reduced expression levels in NSCLC. This reduced expression was associated with worse postoperative survival. The first study to investigate

the large-scale expression of miRNAs in lung cancer identified the upregulation of 35 miRNAs (e.g., miR-21-5p, miR-17-5p, miR-191-5p) and downregulation of three miRNAs (miR-126-5p, miR-30d-5p, and miR-30a-5p), several of which were also dysregulated in several additional solid cancer types (Volinia et al., 2006). Another landmark study identified 43 miRNAs that are dysregulated in NSCLC tissues and found an association between postoperative survival and high miR-155-5p expression (Yanaihara et al., 2006). This study also reported differences between the miRNA profiles of SCC and AD.

Subsequent array- and massively parallel sequencing-based miRNA profiling studies have identified many more miRNAs that are dysregulated in NSCLC. However, the reported sets of potentially relevant miRNAs are different and sometimes conflicting. This issue is addressed in Chapter 4.2 of the experimental part of this thesis.

Exposure to tobacco smoke changes the profile of miRNAs in the lungs of mice and rats (Izzotti et al., 2009a, 2009b), suggesting that miRNA dysregulation is an early event in the pathogenesis of NSCLC and may mediate the subsequent tumorigenic processes. Several miRNAs that are frequently dysregulated in NSCLC have been demonstrated to act as oncomiRs or tumor suppressors, regulating cell cycle, angiogenesis, or epithelial-to-mesenchymal transition. For example, one of the most upregulated miRNAs in NSCLC, miR-21-5p, targets the tumor suppressor *PTEN* (Zhang et al., 2010). The tumor suppressor miR-126-5p, which is frequently downregulated in NSCLC, targets angiogenesis-associated genes *VEGF-A* (Liu et al., 2009) and *EGFL7* (Sun et al., 2010). Although there are numerous examples of validated cancer-associated targets of miRNAs that are dysregulated in NSCLC, their overall effect on tumorigenic processes is still poorly understood because of large number of potential targets, small number of validated targets, and multifunctional nature of individual miRNAs.

The miRNA profile in plasma, serum, or whole blood can be used to differentiate NSCLC patients from healthy controls, identify the metastatic state, and predict disease prognosis or treatment outcome (for comprehensive review; (Del Vescovo et al., 2014)). As an recent example, a serum profile of four miRNAs (miR-193b-5p, miR-301-5p, miR-141-5p, and miR-200b-5p) was found to differentiate NSCLC patients from cancer-free controls with high accuracy (AUC = 0.99) (Nadal et al., 2015). Another recent study identified the serum profile of five miRNAs (miR-483-5p, miR-193a-3p, miR-25-3p, miR-214-3p, and miR-7-5p), which had an AUC of 0.82 in the validation cohort (Wang et al., 2015). However, diagnostic and prognostic biomarker profiles differ substantially between studies, probably due to the variability in technical and analytical methods. MicroRNAs are also detectable and dysregulated in the sputum of NSCLC patients (Xie et al., 2010), adding an additional strategy for how miRNAs can be used as noninvasive biomarkers.

1.4. Genomic variants influencing the function of miRNAs

During the last decade, a large number of trait- and disease-associated genetic polymorphisms have been identified by GWASs. However, only a small proportion of those variants are located in CDSs, suggesting that these variants mainly affect gene regulation rather than the structure of encoded proteins (Hindorff et al., 2009). Indeed, many SNPs influence the expression of nearby (*cis*) and distant (*trans*) genes (Westra et al., 2013). These expression quantitative trait loci (eQTLs) are regulatory loci that can potentially converge their effects to the formation of a trait or disease.

Beside the effects on transcriptional control via regulatory regions in DNA (e.g., transcription factor binding sites and enhancers), methylation patterns, and chromatin packing, noncoding genomic variants could affect the formation of complex traits or diseases through miRNA-mediated regulatory networks. There are differences between human populations in the frequencies of genomic variants in pre-miRNA sequences (Rawlings-Goss et al., 2014) and the expression levels of several mature miRNAs (Huang et al., 2011), raising the possibility that miRNA regulation may play role in the differential susceptibility of diseases (Rawlings-Goss et al., 2014) or other complex traits in populations. Bulik-Sullivan with coworkers prioritized several miRNA-related SNPs that were previously associated with complex traits (Bulik-Sullivan et al., 2013).

SNPs can influence miRNA-meditated regulation via several mechanisms. First, SNPs in promoter or regulatory regions of an miRNA-encoding gene can change the expression of the corresponding miRNA and, thereby, influence its regulatory potential. For example, a SNP in the promoter of miR-146a-5p influences its expression level and is associated with systemic lupus erythematosus (Luo et al., 2011). This mechanism is supported by observations that several SNPs influence the expression of nearby miRNAs (*cis*-miR-eQTLs) (Borel et al., 2011; Huan et al., 2015; Lappalainen et al., 2013).

Second, SNPs or mutations may influence the structure or expression of members of the miRNA processing machinery. Effects of those SNPs may influence the whole miRNome. For example, rs11077 in *XPO5* (encoding exportin 5) has been associated with the severity of several cancers (Campayo et al., 2011; Liu et al., 2014; Ye et al., 2008). Interestingly, a mutation in this gene has been shown to entrap miRNAs in the nucleus of cancer cells (Melo et al., 2010).

Third, SNPs that change the sequence of pri- or pre-miRNAs may have downstream effects on the expression of the mature miRNA. For example, a SNP in the tumor suppressor pri-let-7e led to a reduction of the mature miRNA level in an in vitro system (Wu et al., 2008).

Fourth, SNPs located in the mature miRNA (especially in the seed region) may cause changes in the targetome of the corresponding miRNA by disrupting or creating complementarity with targets. Comprehensive overlap between known genomic variants and miRNA seed regions identified 227 SNPs that

would cause differences in binding with in silico-predicted targets (Gong et al., 2015).

Finally, SNPs in target genes may cause changes in miRNA-mRNA interactions. These SNPs are likely to be found in the 3' UTR of the target gene, as it is the most relevant region for miRNA binding in animals. SNPs in this region may affect the regulation of target genes by miRNAs in multiple ways (Fig. 5). A minor allele of a 3'-UTR SNP may directly affect the binding site by disrupting the functional MRE, creating a new one, changing the strength of the existing MRE, or replacing the MRE of one miRNA with another (SNPs termed as MRE-SNPs). SNPs may also change the secondary structure of the 3' UTR and, therefore, affect its accessibility for RISC. SNPs in splice sites or polyadenylation signals may cause alternative splicing or polyadenylation, changing the length of the 3' UTR and causing loss of multiple MREs. In addition to miRNA-associated mechanisms, SNPs in 3' UTRs may affect regulation by several RNA-binding proteins. Although there is currently no comprehensive study investigating all of these mechanisms, bioinformatics analyses have suggested that most 3'-UTR SNPs influence MREs (64%), whereas a minority affect 3'-UTR folding (24%) or splicing (12%) (Arnold et al., 2012).

Several genomic variants in MREs have been associated with phenotypic traits and pathological conditions. The first study to associate genomic polymorphisms with miRNA action was conducted in papillary thyroid carcinoma (He et al., 2005a). Two SNPs, predicted to change the binding of three miRNAs upregulated in this cancer type, were found in the cancer-associated gene *KIT*. Another early study identified a point mutation in the 3' UTR of *SLITRK1*, which was predicted to enhance the binding of miR-24-1-5p (previously known as miR-189). This mutation was found in two nonrelated Tourette syndrome patients and the results from a luciferase reporter assay supported the functional impact on miRNA binding. The authors concluded that this mutation may have a role in the pathogenesis of this disease (Abelson et al., 2005). Probably the most striking example is a point mutation in the 3' UTR of the gene encoding myostatin (*GDF8*). This mutation was shown to create a potential binding site for muscle-expressed miR-206 and miR-1-3p in the Texel breed of sheep (Clop et al., 2006). Resulting downregulation of *GDF8* causes a specific muscular phenotype for this breed.

After early studies associating genetic variation in MREs with diseases and phenotypes demonstrated the biological relevance of this phenomenon, efforts were made to screen SNPs in MREs at the genome-wide scale, using in silico target prediction algorithms. These studies resulted in several publicly available catalogues of predicted MRE-SNPs, such as polymiRTS (Bao et al., 2007; Bhattacharya et al., 2014), Patrocles (Hiard et al., 2010), miRSNP (Liu et al., 2012), miRNASNP (Gong et al., 2012, 2015), and miRSNPScore (Thomas et al., 2011). Each of these databases uses a different prediction algorithm. Additionally, web tools for the real-time prediction of the effects of MRE-associated variants have been developed (Barenboim et al., 2010; Deveci et al., 2014).

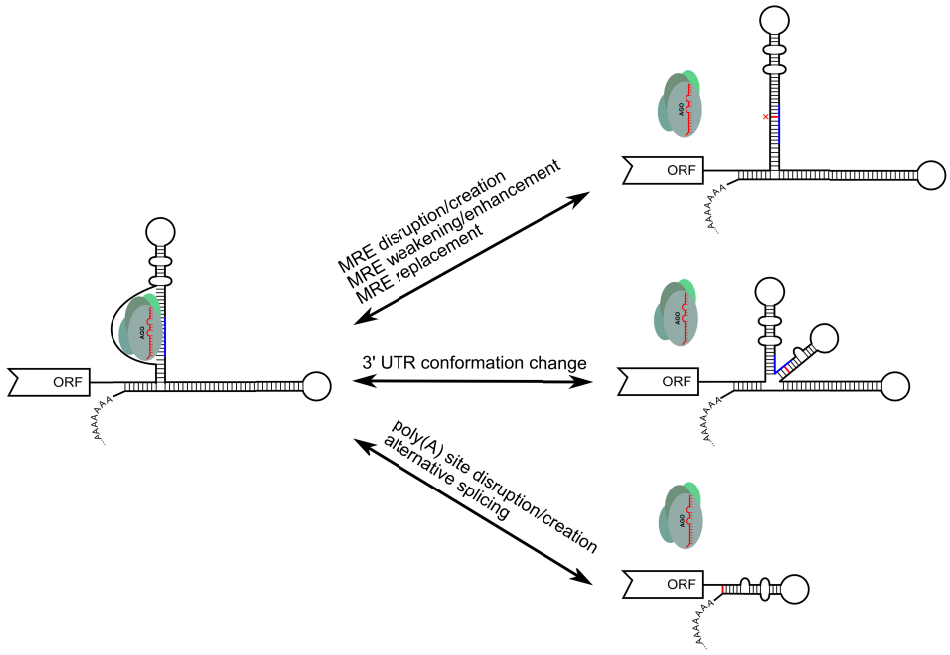


Figure 5. Mechanisms of miRNA-mediated 3'-UTR SNP effects on gene expression.

Because the accuracy of in silico prediction algorithms is not perfect, several additional data layers have been used to prioritize putative MRE-SNPs. For example, Arnold et al. used available CLIP-seq data to enhance the precision of target prediction and identified 37 unique MRE-SNPs (Arnold et al., 2012). Some studies have used experimental support to MRE as an additional filter (Duan et al., 2009; Saunders et al., 2007) or used the requirement of miRNA-target coexpression in the tissue of interest (Saunders et al., 2007; Vaishnavi et al., 2014; Zhang et al., 2012a).

Because miRNA targeting impacts the expression level of the target mRNA, information about *cis*-eQTLs can be used to support the functional effects of the MRE-SNP. For example, Richardson and colleagues used small eQTL datasets (78–83 samples) from the MuTHER study to support the in silico identification of 4 out of 11 putative disease-associated MRE-SNPs (Richardson et al., 2011). Another study used data from HapMap 3 LCL samples and identified 130 MRE-SNPs that had a *cis*-eQTL effect on target gene expression (false discovery rate [FDR] < 0.001). The same study hypothesized that, in addition of "fine-tuning" the targetome, miRNA targeting may enhance variations in the expression of some target genes (Lu and Clark, 2012).

2. AIMS OF THE STUDY

The aims of the present doctoral thesis were as follows:

1 To identify differentially expressed miRNAs in NSCLC, determine their effects on cellular signaling pathways, and find miRNAs with potential diagnostic and prognostic value.
2 To determine a consensus set of differentially expressed miRNAs in lung cancer by meta-analyzing the results of previously published miRNA profiling studies.
3 To identify genetic variants that alter the expression levels of host genes by miRNA-mediated mechanisms.

3. RESULTS AND DISCUSSION

3.1. MicroRNA expression changes in Estonian NSCLC cohort (Ref I)

Lung cancer is the second most common cancer type in Estonia and the main cause of cancer-related deaths (Estonian Cancer Registry, 2008-2012). Although the survival rate of patients with NSCLC is generally poor, a much better prognosis is possible when the disease is discovered in the early stages. Different treatment decisions can be made depending on the precise molecular subtype of the disease, highlighting the need for more accurate diagnostic and prognostic biomarkers.

This part of the thesis describes the identification of miRNA expression aberrations occurring in the early stages of NSCLC development by miRNA expression profiling analysis in a cohort of Estonian patients with early-stage NSCLC. Identified miRNAs are expected to have a higher likelihood of having a functional effect in cancer pathogenesis compared to later-stage "passenger" aberrations. Additional analyses were performed to evaluate miRNA-associated effects on cellular signaling pathways and to identify miRNAs associated with patient survival.

3.1.1. Cohort used in this study

Samples in this study were collected from NSCLC patients who underwent curative resection at Tartu University Lung Hospital between 2002 and 2008. The study was approved by the Ethics Committee on Human Research of Tartu University and all enrolled patients signed a written informed consent form before participation. Biological material was collected by Dr. Tõnu Vooder in a standardized way (Ref I for details). Histological examination for tissue samples was made by Dr. Retlav Roosipuu. Clinical and pathological data were also collected.

The present study incorporated a subset of 38 early-stage NSCLC and 27 adjacent cancer-free lung samples (including 24 paired samples). Samples were selected from patients who had received neither preoperative irradiation nor chemotherapy because such treatment may have an independent effect on miRNA expression (Simone et al., 2009). Six patients received adjuvant chemotherapy after resection and this fact was taken into account in subsequent survival analyses. The 28 NSCLC samples included 18 SCC and 20 AD samples. Thus, the results from this study reflect expression changes in NSCLC in general, rather than in one specific subtype. The overall cohort characteristics are summarized in Table 1.

Table 1. Clinical and pathological characteristics of the samples.

Characteristic	SCC[a]	AD[b]	NC[c]
Number of samples	18	20	27
Sex			
Male	18	14	23
Female	0	6	4
Stage			
Ia	2	12	7
Ib	14	7	17
IIa	0	0	0
IIb	2	1	3
Smoking			
Yes	17	15	24
No	0	4	3
unknown	1	1	
Age			
≤ 63 y	7	9	11
> 63 y	11	11	16
Range	44-79	41-79	41-77
Median	67	64	67
Survival status			
Alive	9	16	20
Deceased	9	4	7

[a]SCC, squamous cell carcinoma, [b]AD, adenocarcinoma, [c]NC, cancer-free tissue adjacent to tumor from NSCLC patients. Disease stage of corresponding individuals is shown.

3.1.2. MicroRNAs dysregulated in Estonian lung cancer cohort

Illumina miRNA BeadArray expression arrays were used to profile the expression of 858 miRNAs (miRBase v14). At the time of the study, this platform provided the most comprehensive coverage of the human miRNome. Standard data preprocessing and statistical analyses were performed with the LIMMA package (Linear Models for Microarray Analysis; (Ritchie et al., 2015; Smyth, 2004)). This study identified 39 up- and 33 downregulated miRNAs in NSCLC compared to the cancer-free samples (FDR < 0.01). When a stricter threshold for expression change (FDR < 0.01, fold change [FC] > 2; Fig. 6) was applied, 31 up- and 29 downregulated miRNAs remained. The most significantly dysregulated miRNAs are presented in Table 2 and the full list can be found in the Supplementary Material of Ref I.

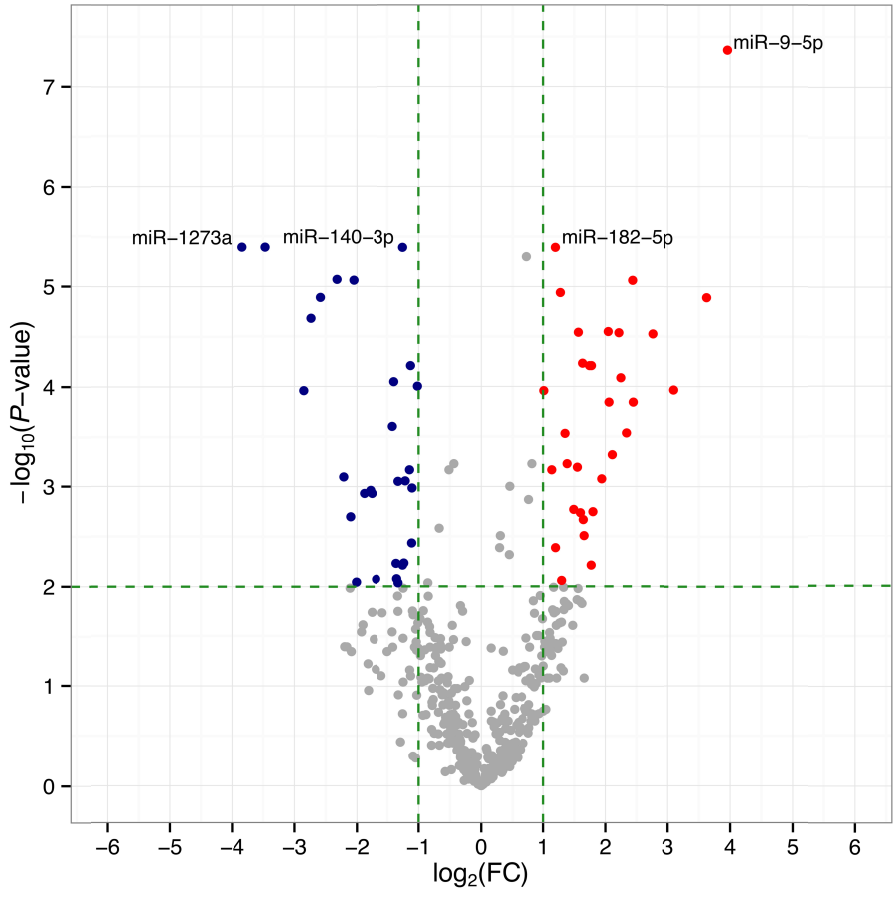


Figure 6. Volcano plot depicting results of differential miRNA expression analysis. X-axis represents expression-level difference (fold change, FC) on the $\log_2$ scale. Y-axis represents level of statistical significance [$-\log_{10}$(FDR-corrected *P*-value)]. Vertical dashed lines show the two-fold expression change as a cutoff. Horizontal dashed line indicates the threshold of statistical significance (FDR = 0.01). Red points depict up- and blue points downregulated miRNAs.

Many of the identified miRNAs are known to be associated with lung cancer. The most upregulated miRNA in the dataset, miR-9-5p (FDR = 4.3×10^{-8}, FC = 15.5), has been associated with lung cancer (Crawford et al., 2009; Volinia et al., 2006; Xu et al., 2014; Yanaihara et al., 2006) and several other cancers (Cai and Cai, 2014; Lehmann et al., 2008; Wu et al., 2014). Although this miRNA showed strong and unidirectional cancer-associated upregulation in the dataset (Fig. 4), other studies have reported both up- (Crawford et al., 2009; Volinia et al., 2006; Xu et al., 2014) and downregulation (Yanaihara et al., 2006) in lung cancer.

Table 2. Thirty most significant differentially expressed miRNAs in NSCLC.

	miRNA	*P*-value*	FC[a]	Host gene(s)[b]	Location
Upregulated				*C1orf61* [+]	1q22
	miR-9-5p	4.30×10^{-8}	15.5	*LOC64532* [+]	5q14.3
				intergenic	15q26.1
	miR-182-5p	4.06×10^{-6}	2.3	intergenic	7q32.2
	miR-200a-5p	8.67×10^{-6}	5.4	intergenic	1p36.33
	miR-151:9.1[#]	1.15×10^{-5}	2.4	*PTK2* [+]	8q24.3
	miR-205-5p	1.29×10^{-5}	12.3	*LOC642587* [+]	1q32.2
	miR-183-5p	2.82×10^{-5}	4.1	intergenic	7q32.2
	miR-130b-5p	2.87×10^{-5}	3.0	intergenic	22q11.21
	miR-149-5p	2.90×10^{-5}	4.6	*GPC1* [+]	2q37.3
	miR-193b-3p	2.98×10^{-5}	6.8	intergenic	16p13.12
	miR-339-5p	5.86×10^{-5}	3.1	*C7orf50* [+]	7p22.3
	miR-196b-5p	6.19×10^{-5}	3.4	intergenic	7p15.2
	miR-224-5p	6.19×10^{-5}	3.4	*GABRE* [+]	Xq28
	miR-31-5p	8.20×10^{-5}	4.8	*LOC554202* [+]	9p21.3
	miR-196a-5p	1.08×10^{-4}	8.6	intergenic	17q21.32
				intergenic	12q13.13
	miR-423-3p	1.10×10^{-4}	2.0	*CCDC55* [+]	17q11.2
	miR-708-5p	1.44×10^{-4}	5.5	*ODZ4* [+]	11q14.1
	miR-106b-3p	1.44×10^{-4}	4.2	*MCM7* [+]	7q22.1
	miR-210-3p	2.92×10^{-4}	5.1	intergenic	11p15.5
Downregulated	miR-1273a	4.06×10^{-6}	14.3	*RGS22* [+]	8q22.2
	miR-206	4.06×10^{-6}	11.1	intergenic	6p12.2
	miR-140-3p	4.06×10^{-6}	2.4	*WWP2* [+]	16q22.1
	miR-338-3p	8.51×10^{-6}	5.0	*AATK* [+]	17q25.3
	miR-101-3p	8.67×10^{-6}	4.1	intergenic	1p31.3
				RCL1 [+]	9p24.1
	miR-144:9.1[#]	1.29×10^{-5}	6.0	intergenic	17q11.2
	miR-1285-3p	2.08×10^{-5}	6.7	*KRIT1*	17q11.2
	miR-130a-3p	6.19×10^{-5}	2.2	intergenic	11q12.1
	miR-486-5p	8.97×10^{-5}	2.6	*ANK1* [+]	8p11.21
	miR-24-2-5p	9.94×10^{-5}	2.0	intergenic	19p13.12
	miR-144-5p	1.10×10^{-4}	7.2	intergenic	7q21.2
	miR-30a-5p	2.51×10^{-4}	2.7	*C6orf155*	6q13

[*] *P*-values are corrected by Benjamini–Hochberg method. [a]FC, fold change; [b][+] miRNA and host gene are transcribed in the same direction; [#]9.1, given probe is specific to miRBase v9.1 version of the miRNA.

A previous study showed that miR-205-5p (FDR = 1.29×10^{-5}, FC = 12.3) is highly expressed in SCC and capable of distinguishing SCC from AD (Lebanony et al., 2009). Although the present dataset consisted of SCC and AD samples, upregulation was present in most patients (Fig. 7). One of the experimentally supported targets of this miRNA is the well-known tumor suppressor

PTEN (Sun et al., 2013). The most strongly downregulated miRNA, miR-1273a, has not been investigated before. However, several of the downregulated miRNAs have been implicated in lung and other cancer types. For example, miR-206 was recently confirmed to be downregulated in NSCLC, and transcription factor *SOX9* was experimentally identified as its target (Zhang et al., 2015b). Similar studies have identified additional potentially relevant targets, like cancer-associated *MET* and *BCL2* (Chen et al., 2015; Sun et al., 2015).

3.1.3. Pairwise expression patterns

In order to serve as diagnostic biomarkers or drug targets, miRNAs should show consistent changes in expression in tumors. Therefore, this study used the subset of array expression data containing paired samples of cancerous and non-cancerous lung tissue (n = 24) to visualize expression changes of the most differentially expressed miRNAs between samples from the same individual (Fig. 7).

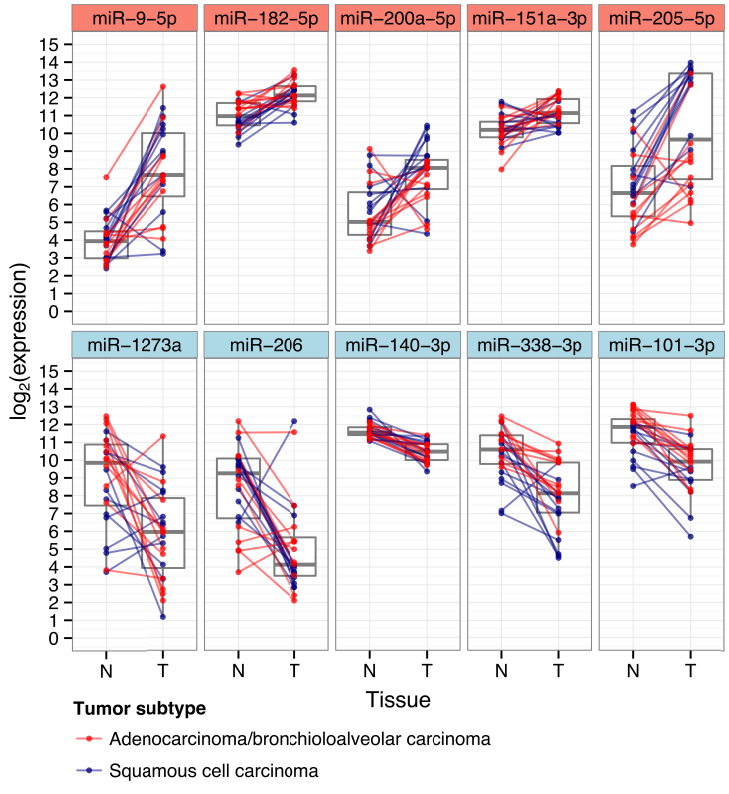


Figure 7. Pairwise expression pattern of the five most strongly up- and downregulated miRNAs. Each point represents an individual tissue sample. Lines connect samples from the same individual. N: cancer-free lung sample; T: tumor sample. Up- and downregulated miRNAs are in the top and bottom rows, respectively. Cancer subtypes are represented with different colors.

Although expression changes were generally consistent between sample pairs, there were some miRNAs with incoherent patterns. Some highly significant miRNAs with large overall changes in expression had inconsistent sample pairs (e.g., miR-1273a), whereas other miRNAs with more modest expression changes showed remarkable consistency (e.g., miR-140-3p). This finding may indicate that an miRNA does not necessarily need a large change in expression to serve as a potential diagnostic biomarker. As one miRNA may target hundreds of genes and regulate several pathways, modest but consistent changes in expression may point to possible oncomiR or tumor-suppressor miRNAs.

3.1.4. Validation of miRNA array data

Next, qRT-PCR was used to validate the miRNA array results. First, the subset of eight pairs of samples was used to validate changes in expression of four miRNAs that were upregulated in NSCLC (miR-9-5p, miR-196a-5p, miR-149-5p, and miR-205-5p; Fig. 8). Secondly, eight sample pairs from the NSCLC cohort that had not been profiled with Illumina arrays were used to validate the changes in expression for 10 aberrantly expressed miRNAs in the independent sample set. Among the investigated miRNAs, all but one (miR-1273a) were detectable in that sample set. In most cases, the direction of expression change was in concordance with array data (Fig. 8). However, the magnitude of change varied between array data and qRT-PCR normalization methods.

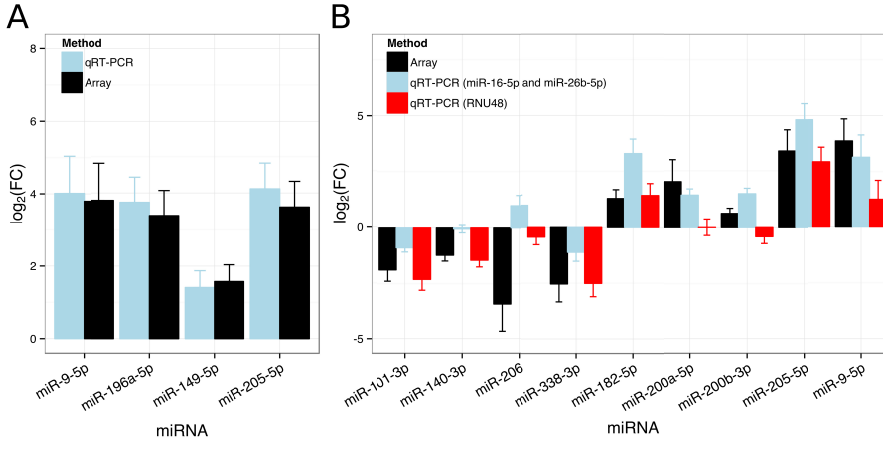


Figure 8. Validation of miRNA expression. A. Validation of microarray data by qRT-PCR. Bars represent average $\log_2$(FC) calculated for eight sample pairs on the microarray. Error bars indicate standard error of the mean (SEM). Geometric mean of the expression levels of miR-16-5p and miR-26b-5p was used as an endogenous reference for qRT-PCR. B. Validation of miRNA expression in eight additional sample pairs not profiled by the microarray. Bars represent $\log_2$(FC) and error bars indicate SEM. Results using two normalization methods (geometric mean of miR-16-5p and miR-26b-5p or expression of RNU48) are shown.

Two different endogenous references were used to normalize the qRT-PCR data: small nucleolar RNA 48 (RNU48) and the geometric mean of two highly expressed miRNAs (miR-16-5p and miR-26b-5p). Those genes were chosen because they have been prioritized as suitable references for similar analyses in a breast cancer dataset (Davoren et al., 2008).

3.1.5. Association between miRNA expression and patient survival

The next step was to identify whether miRNA expression was associated with patient survival. For this analysis, the data subset consisting only of tumor samples was used. Two samples were omitted due to the short survival time (<15 days) after resection because of the possibility that death was not directly related to the cancer. Therefore, the final analysis included 36 NSCLC samples. A Cox regression model was fitted on expression data of each miRNA. A permutation test was performed by assigning the survival time and status randomly 10,000 times. Using the permuted *P*-value threshold of 0.01, eight miRNAs had significant associations with survival time.

Patient survival may be influenced by several factors aside from possible cancer-related miRNAs. Therefore, the identified miRNAs were thoroughly analyzed to determine whether the observed association was an independent effect. Multivariate Cox regression was used to correct for the effects of possible covariates: smoking status (in pack-years), histological type (SCC or AD), tumor stage, postoperative chemotherapy, age, and sex. One miRNA, miR-374a-5p, had an independent association with patient survival (hazard ratio [HR] = 0.353, $P = 0.008$). High- and low-expression groups were defined on the basis of the median miR-374a-5p expression level, and survival of those groups was compared by Kaplan–Meier survival curves and log-rank tests. The high-expression group showed significantly longer survival times compared to the low-expression group ($P = 0.018$, log-rank test; Fig. 9).

The miRNA miR-374a-5p is upregulated in several cancer types (He et al., 2015; Xu et al., 2015) and mediates the upregulation of Wnt/β-catenin signaling in breast cancer cells (Cai et al., 2013). Wang et al. found that this miRNA is involved in greater gefitinib resistance in lung cancer and that this miRNA is targeting *WNT5A* (Wang et al., 2014). However, the same study demonstrated the opposite effect of miR-374a-5p expression with lung cancer patient disease-free survival compared to the pattern observed here (Wang et al., 2014). Reasons for this discrepancy are unclear, but may reflect differences arising from distinct ethnic backgrounds, variability in cohort characteristics (e.g., most NSCLC patients in the mentioned study had stage III or IV cancer, compared to early-stage patients used in the current study), or small sample sizes.

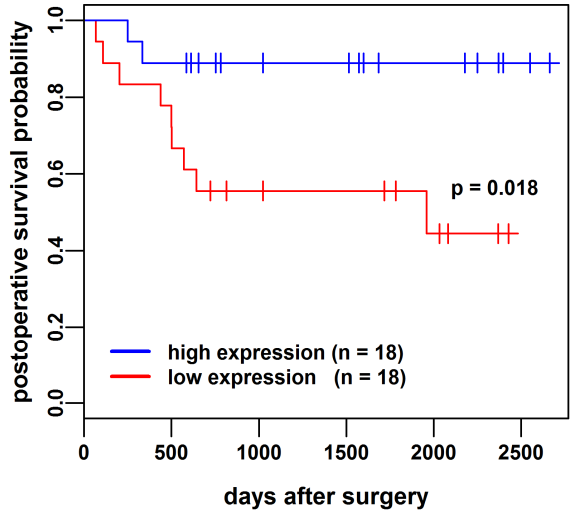


Figure 9. miR-374a-5p association with patient postoperative survival. Kaplan–Meier plot showing postoperative survival of NSCLC patients depending on the miR-374a-5p expression level. Log-rank test *P*-value is shown.

3.1.6. Regulatory pathways influenced by differentially expressed miRNAs

3.1.6.1. Target prediction

Putative targets for most dysregulated miRNAs were predicted by using the miRNA target prediction algorithms Pictar 4-way, PITA, TargetScan 5.1, miRanda, and DIANA-microT 3.0. To obtain a reliable set of target genes, a conservative strategy was used. The gene was considered to be the target of the corresponding miRNA if at least three of the five target prediction algorithms supported the prediction.

3.1.6.2. Enrichment analyses

To gain insight into the biological functions of the most significantly dysregulated miRNAs in NSCLC, gene set enrichment analysis was conducted by using annotations from the Kyoto Encyclopedia of Genes and Genomes (KEGG). Enriched pathways for the 10 most up- and downregulated miRNAs were largely overlapping (Fig. 10), being linked to cell mobility, differentiation, and proliferation (“Axon guidance”, “Actin cytoskeleton”, “Focal adhesion”, etc.). Several cancer-associated pathways were also enriched (“Pathways in cancer”, “Prostate cancer”, etc.). Separate pathway enrichment analysis for targets of miR-374a suggested that this miRNA is involved in tumorigenesis-related processes (“Pathways in cancer”, FDR = 0.024).

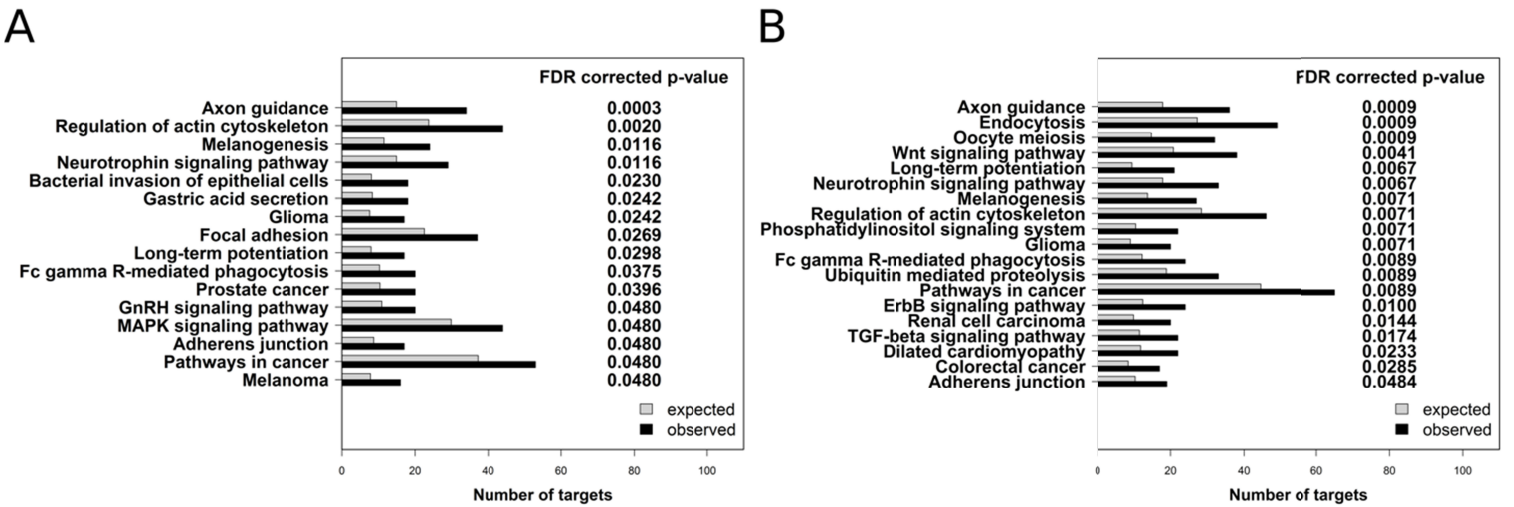


Figure 10. KEGG pathway enrichment of potential targets of the 10 most up- (A) and downregulated (B) miRNAs. Shown are expected and observed gene counts and FDR-corrected *P*-values.

For comparison, the enrichment analyses were made with the results of a previous mRNA profiling study that used individuals from the same Estonian NSCLC cohort (Välk et al., 2010). Pathways enriched among upregulated genes were involved in cell proliferation ("p53 signaling", "cell cycle", "DNA replication"). Downregulated genes were enriched by immune-related pathways.

3.1.7. Concordance between miRNA and mRNA expressions

Next, the research addressed whether dysregulation on the miRNA level was responsible for the differential expression of some proportion of mRNAs. Enrichment among the dysregulated mRNAs (Välk et al., 2010) was tested by using the in silico-predicted targets of the 10 most up- and downregulated miRNAs, followed by assessment of whether the enrichment was consistent with the logic of miRNA regulation. Genes showing downregulation in the gene expression study were enriched by the targets of the 10 most upregulated miRNAs (Fisher exact test, $P = 0.021$, odds ratio [OR] = 1.37, 95% confidence interval [CI] = 1.04–1.78). This finding was in agreement with the logic of miRNA regulation and may suggest possible coordinated regulation among some of the cancer-associated miRNA-target pairs in NSCLC. Interestingly, targets of the 10 most downregulated miRNAs were underrepresented among the upregulated mRNAs (Fisher exact test, $P = 0.0004$, OR = 0.49, 95% CI = 0.30–0.75). On the other hand, there was no significant association between the two downregulated lists (predicted targets of downregulated miRNAs and downregulated mRNAs) and the two upregulated lists.

3.2. Meta-analysis of differentially expressed miRNAs in lung cancer (Ref II)

Although the results were consistent with previous knowledge for many differentially expressed miRNAs (e.g., miR-205-5p, miR-182-5p, miR-30a-5p, etc.), differences with published miRNA expression studies were also observed. For example, miR-9-5p, the most upregulated miRNA in the Estonian NSCLC cohort, was reported to be upregulated in only two research papers and downregulated in a landmark study (Yanaihara et al., 2006). The most strongly downregulated miRNA (miR-1273a), as well as several other significantly dysregulated miRNAs, were not previously associated with lung cancer. Discrepancies between the results of previously published individual studies were also frequently observed, frustrating attempts to conclude which miRNAs are actually cancer-associated and which represent technical or analytical artifacts. These observations indicated the need for a comprehensive and systematic meta-analysis, which would yield a more credible and robust set of differentially expressed miRNAs, which, in turn, could serve as candidates in

subsequent research for the identification of drug targets or diagnostic biomarkers.

The preferred method of meta-analysis of gene expression studies involves complete reanalysis of raw expression data. However, this kind of rigorous strategy is often not possible due to the unavailability of raw expression data and technical obstacles (e.g., differences between studies in miRNA expression platforms, preprocessing steps, and numbers of detectable miRNAs available at the time of the study), which make the integration of raw miRNA expression datasets very complicated. Therefore, a rank-based approach was used to summarize and meta-analyze the results of previously published lung cancer miRNA profiling studies. Additional bioinformatics approaches were used to evaluate the effects of identified miRNAs on cellular signaling pathways.

3.2.1. Selection and preprocessing of NSCLC miRNA expression profiling datasets

Comprehensive searches in the Scopus database were conducted by using the following search term: TITLE-ABS-KEY((mirna* OR microrna* OR mir-*) AND profil* AND lung AND (cancer* OR tumor* OR tumour*)) in September 2012. Additional searches were conducted in the Gene Expression Omnibus (GEO) and ArrayExpress databases, which consist of publicly available expression datasets. The search was limited to original English-language studies of miRNA expression changes between cancerous and noncancerous lung tissues. Studies analyzing preselected candidate miRNAs or comparing different groups of cancerous tissues were removed to minimize the sources of bias in the subsequent meta-analysis.

Using those criteria, 21 lists of differentially expressed miRNAs from 20 studies were identified and enrolled (Table 3). Studies utilized several expression profiling platforms detecting different numbers of miRNAs (mean: 560; range: 258–858 miRNAs) from different miRBase versions. Some miRNA lists contained viral miRNAs or probes detecting pre-miRNA sequences. Thus, the assembly of individual miRNA lists involved removal of nonhuman miRNAs and standardization of miRNA names to a common miRBase version (v19) before meta-analysis.

Table 3. Characteristics of analyzed datasets.

Ref.	Acronym	Region	Assay type	Number of miRNA probes	Tumor type	Number of samples
Volinia et al. 2006	VL	North America	Custom microarray	228	N/S	123 pairs
Yanaihara et al. 2006	YH	North America	Custom microarray	352	SCC, AD	104 pairs
Cho, Chow, and Au 2009	CH	Asia	Agilent Human miRNA Array	470	AD	10 pairs
Crawford et al. 2009	CR	North America	qPCR, Mega Plex kit	500	AD	8 pairs + 4 TU
Raponi et al. 2009	RA	North America	Ambion mirVana Bioarray v2	328	SCC	61 TU, 10 N
Seike et al. 2009	SE	North America	Custom microarray	389	SCC, AD	28 pairs
Son et al. 2009	SO	Asia	Genomic Tree Inc.	318	SCC, AD	7 pairs
Gao et al. 2010	G10	Asia	Exiqon miRCURY LNA array	730	SCC, AD	8 pairs
Xing et al. 2010	XI	North America	Affymetrix GeneChipR array	818	SCC	15 pairs
Yang et al. 2010	YG	Asia	Exiqon miRCURY LNA array	711	SCC	3 pairs
Yu et al. 2010	YU	North America	System Biosciences, Taqman	377	AD	20 pairs
Boeri et al. 2011	BO	North America	Custom microarray	235	SCC, AD	24 pairs + 4 TU
Gao et al. 2011	G11	Asia	Exiqon miRCURY LNA array	730	SCC	4 pairs
Lee et al. 2011	LE	Asia	Custom microarray	735	SCC, AD	6 pairs
Ma et al. 2011	MA	Asia	Illumina microarray V2	858	SCC, AD	6 pairs
Nymark et al. 2011	NY	Europe	Agilent Human miRNA v2	723	various	26 pairs
Puisségur et al. 2011	PU	Europe	Custom microarray	409	SCC, AD	20 pairs
Võsa et al. 2011	VS	Europe	Illumina microarray V2	858	SCC, AD	24 pairs + 14 TU + 3 N
Wang et al. 2011	WG	Asia	µParaflo microfluidic chip	564	SCC, AD	23 pairs
Jang et al. 2012	JA	North America	Illumina microarray V2	858	AD	103 pairs

AD, adenocarcinoma; SCC, squamous cell carcinoma; N/S, not specified; TU, tumor samples; N, noncancerous samples; pairs, TU and N samples from the same patient.

Sizes of the analyzed sample groups varied greatly between studies (range: 3–123 sample pairs), indicating differences in the power to detect aberrations. Two SCLC samples were included in the final meta-analysis, due to the specific asbestos exposure-related question of the corresponding study (Nymark et al., 2011). However, most studies used a merged set of SCC and AD samples. Therefore, results of this meta-analysis reflect aberrations general to NSCLC. Altogether, 598 tumor and 528 noncancerous lung samples were used in this meta-analysis.

3.2.2. Cluster analysis of lung cancer miRNA expression profiling datasets

To analyze and visualize similarities between miRNA lists, hierarchical cluster analysis was performed on the ranked lists of miRNAs. Rank matrices were calculated and normalized by using the RobustRankAggreg package (Kolde et al., 2012) and information about miRNAs detectable by each study. Rank matrices from up- and downregulated miRNA lists were merged, and cluster analysis was applied to find differences in the ranked lists (Fig. 11).

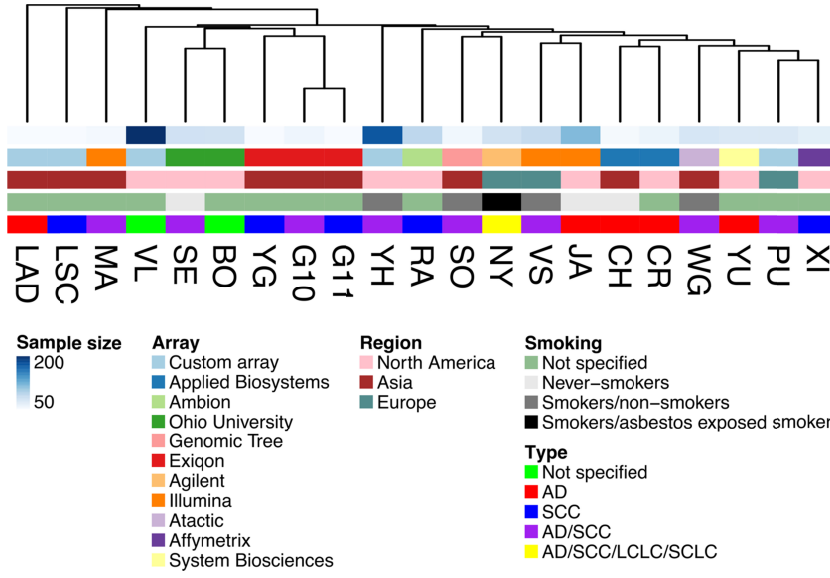


Figure 11. Clustering of lists of differentially expressed miRNAs in lung cancer. Clustering was determined from normalized rank matrices derived from ranked lists of differentially expressed miRNAs. Clustering is based on the Spearman correlation distance measure with average linkage. Full figure with miRNA heatmap is available in the Supplementary Figure 1 of Ref II.

The miRNA profiling platform was the main factor to influence clustering. Two sufficiently powered studies utilizing Illumina miRNA arrays (Jang et al., 2012; Võsa et al., 2011), as well as studies using Exiqon miRCury arrays (Gao et al., 2010b, 2011; Yang et al., 2010) tended to cluster together. The most similar differentially expressed miRNA lists were acquired from two studies from one work group (Gao et al., 2010b, 2011). The clustering may be caused by the use of an identical array or similar laboratory conditions and preprocessing methods. No information was available on whether those two studies shared some of the same samples (8 and 4 sample pairs, respectively) and, according to the study enrollment protocol, treated those as independent studies in the meta-analysis. However, if this were to be the case, then the overlap of used samples may explain the tight clustering.

Several studies had small sample sizes with less than 10 sample pairs of mostly Asian origin (Gao et al., 2010b, 2011; Lee et al., 2011b; Ma et al., 2011b; Son et al., 2009). Three of those studies showed the largest difference from the remaining studies, possibly indicating the requirement of reasonable sample sizes to achieve robust results in miRNA profiling studies.

3.2.3. Metasignature of miRNA expression in lung cancer

To rank the lists of differentially expressed miRNAs acquired from different studies, the up- and downregulated miRNAs were treated separately. *P*-values describing the significance of differential expression or FCs were used when *P*-values were not reported or not informative. For meta-analysis, the RobustRankAggreg algorithm (Kolde et al., 2012) was applied separately for up- and downregulated miRNA lists. Based on the ranking of each element in the aggregated list compared to the null model assuming random ordering, this method assigns a *P*-value for each element. *P*-values were corrected for multiple testing by the Bonferroni method, based on the largest number of miRNA probes available in any of the used platforms (N = 858; Võsa et al. 2011). Using this method, the metasignature of seven up- and eight down-regulated miRNAs was identified (Table 4).

To check whether the significant effect of the miRNA may be driven by one outlier list, one list was randomly removed from the analysis 10,000 times, and the average *P*-value was calculated. Resulting *P*-values indicated that significant results were not artifacts driven by one specific miRNA list.

Although the resulting signature contains several miRNAs that are established oncomiRs or tumor suppressors, it lacks some of the “classical” established oncomiRs (e.g., miR-17-92 family members) and tumor suppressors (e.g., let-7 family members or miR-34a-5p). One possibility is that the expression change of those miRNAs in lung cancer is not very strong and, therefore, the miRNAs did not reach statistical significance in the overall meta-analysis. Dysregulation of those miRNAs may also be specific to certain clinicopathological features (e.g., cancer subtype, stage, etc.), which could cause the signal to be diluted in the meta-analysis.

Table 4. Lung cancer meta-signature miRNAs.

miRNA	Chromo-some	Corrected *P*-value	Permutation *P*-value	Studies	TargetScan seed family	miRNA Cluster	miRBase release*
Upregulated							
miR-21-5p	17q23.1	1.66×10^{-14}	4.69×10^{-16}	13	miR-21/590-5p	-	1 (12/2002)
miR-210-3p	11p15.5	6.04×10^{-11}	7.98×10^{-13}	13	miR-210	-	2 (07/2003)
miR-182-5p	7q32.2	2.83×10^{-8}	4.42×10^{-10}	9	miR-182	mir-182/96/183	3 (01/2004)
miR-31-5p	9p21.3	1.17×10^{-4}	8.83×10^{-7}	8	miR-31	-	1 (12/2002)
miR-200b-3p	1p36.33	1.30×10^{-3}	9.04×10^{-6}	8	miR-200bc/429/548a	mir-200ab/429	2 (07/2003)
miR-205-5p	1q32.2	6.70×10^{-3}	4.31×10^{-5}	7	miR-205	-	3 (01/2004)
miR-183-5p	7q32.2	3.46×10^{-2}	1.57×10^{-4}	7	miR-183	mir-182/96/183	2 (07/2003)
Downregulated							
miR-126-3p	9q34.3	6.74×10^{-12}	1.35×10^{-13}	13	miR-126-3p	-	2 (07/2003)
miR-30a-5p	6q13	1.40×10^{-9}	1.99×10^{-11}	12	miR-30abcdef/30abe-5p/384-5p	mir-30a/c-2	1 (12/2002)
miR-30d-5p	8q24.22	2.00×10^{-8}	3.55×10^{-10}	9	miR-30abcdef/30abe-5p/384-5p	mir-30b/d	1 (12/2002)
miR-486-5p	8p11.21	4.36×10^{-7}	6.87×10^{-9}	8	miR-486-5p/3107	-	8 (02/2006)
miR-451a	17q11.2	7.36×10^{-5}	4.85×10^{-7}	9	miR-451	mir-4732/144/451	7 (06/2005)
miR-126-5p	9q34.3	9.37×10^{-5}	7.27×10^{-7}	8		-	2 (07/2003)
miR-143-3p	5q32	6.64×10^{-4}	4.96×10^{-6}	8	miR-143/1721/4770	mir-143/145	3 (01/2004)
miR-145-5p	5q32	1.08×10^{-3}	8.70×10^{-6}	8	miR-145	mir-143/145	2 (07/2003)

Corrected *P*-values are adjusted by Bonferroni correction for the largest number of detectable probes over platforms (858). Permutation *P*-values are averaged after leaving out one random miRNA list for 10,000 times.

In three datasets, miRNA profiling was performed in never-smoker lung cancer patients (Cho et al., 2011; Jang et al., 2012; Seike et al., 2009). Because smoking is a major contributor to lung cancer risk, it is possible that the molecular mechanisms leading to disease differ between those NSCLC groups. To ascertain if we could identify a never-smokers–specific miRNA signature, the three never-smoking miRNA lists were used in a separate meta-analysis. Unfortunately, no statistically significantly dysregulated miRNAs were identified, possibly due to the small number of original studies. However, several meta-signature miRNAs were present in individual never-smoker lists, including miR-210-5p, miR-183-5p, and miR-126-5p.

Meta-analysis of five SCC- and AD-specific lists resulted in the identification of a small number of miRNAs dysregulated in specific NSCLC subtypes. For SCC, the only significantly upregulated miRNA was miR-205-5p (Bonferroni-corrected $P = 0.005$), whereas miR-30a-5p and miR-126-5p showed significance in downregulated lists (Bonferroni-corrected $P = 0.0002$ and 0.0009, respectively). For AD-specific lists, only upregulated miR-182-5p, miR-21-5p, and miR-210-5p were identified (Bonferroni-corrected $P = 0.0004$, 0.0011, and 0.0062, respectively). The upregulation of miR-205-5p in only SCC is in concordance with previous knowledge that upregulation of this miRNA is stronger in SCC (Fig. 7), and that this miRNA can be used for distinguishing between SCC and AD samples with high sensitivity and specificity (Lebanony et al., 2009; Patnaik et al., 2015).

3.2.4. Regulatory pathways associated with lung cancer miRNA metasignature

3.2.4.1. Target prediction

Three up-to-date target prediction algorithms (TargetScan v6.1, PicTar, and DIANA-microT-CD) were used to identify targets for each metasignature miRNA (settings for each algorithm are in the Methods of Ref II). Two databases of experimental evidence for mRNA-miRNA binding were used: starBase, which uses several CLIP-seq datasets to support in silico predictions; and TarBase v6.0, which was the most comprehensive database of experimentally supported miRNA-mRNA interactions available at the time. To obtain a reliable set of putative targets, only genes that were predicted to be the targets by at least two of the three algorithms and/or that had some experimental support were used as the “high-confidence” set of miRNA targets.

Some of the metasignature miRNAs (e.g., miR-30a-5p, miR-30d-5p, and miR-182-5p) had a large number of targets, indicating their involvement in a broad spectrum of physiological processes, but also their potentially weaker effect on an individual target. Other miRNAs (e.g., miR-126-3p, miR-451a, and miR-210-3p) had a small number of targets, suggesting their involvement in a narrower spectrum of processes.

3.2.4.2. Enrichment analyses

To elucidate the biological functions and pathways connected with the meta-signature, enrichment analyses were conducted for the high-confidence lists of predicted targets of both individual miRNAs and the combined set of all targets. The most significantly enriched pathways were associated with cell signaling (i.e., "EGF receptor signaling pathway," "MAPK signaling pathway," "Wnt signaling pathway"), cell mobility ("Regulation of actin cytoskeleton"), and cancer ("Pathways in cancer"). There was no clear difference in functions between up- and downregulated members of the miRNA metasignature (Fig. 12, Supplemental Fig. 4 of Ref II).

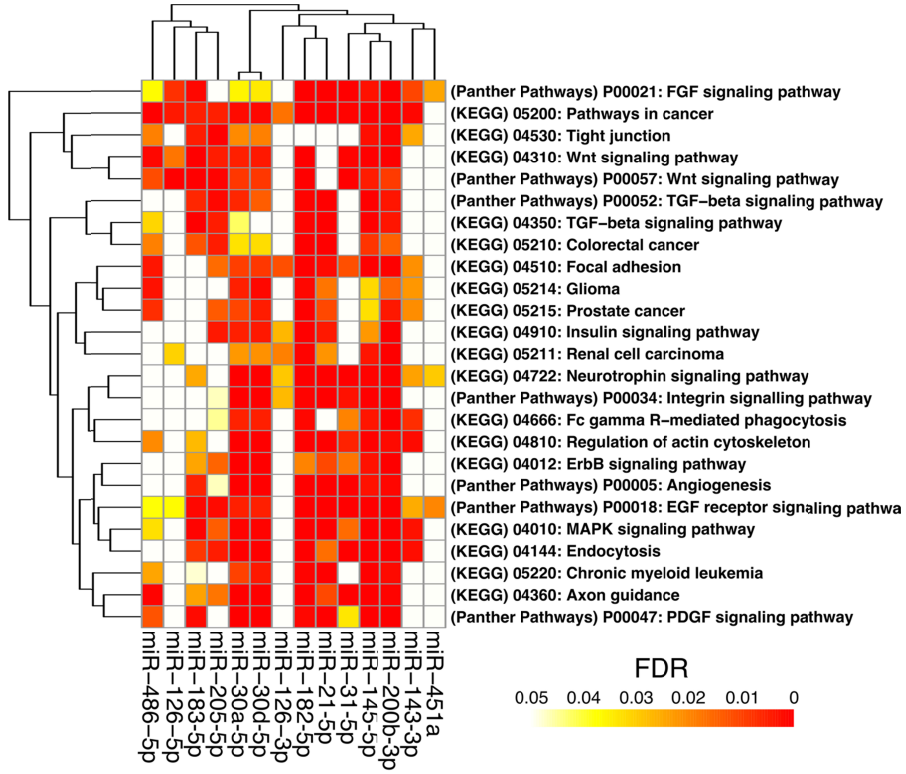


Figure 12. Pathway enrichment of miRNA targets. Consensus targets and results from GeneCodis web tool were used to construct a heatmap showing the results of pathway enrichment analysis. Color intensity represents the FDR-corrected *P*-value. Clustering was performed by using the Pearson correlation distance metric and average linkage method. Only pathways significant for most metasignature miRNAs are shown. Full data are available in Supplementary Fig. 4 of Ref II.

3.3. Effect of genetic variants on miRNA binding (Ref III)

During the last decade, numerous genomic variants have been associated with human traits and diseases, including cancer. However, the molecular mechanisms behind their effects are largely unknown. As most GWAS loci are not located in protein-coding regions, these loci most likely act through regulation of gene expression. A recent large-scale blood eQTL meta-analysis (Westra et al., 2013) identified thousands of *cis*-eQTLs and about 200 *trans*-eQTLs overlapping with established GWAS SNPs. The authors observed the overrepresentation of *trans*-eQTL SNPs in silico-predicted miRNA binding sites, suggesting that genetic variations in miRNA regulatory pathways may influence the formation of traits or risk of diseases. Despite the existence of bioinformatics studies analyzing the global effect of genetic variation on miRNA binding sites, only a handful of these studies integrated eQTL data (Gamazon et al., 2012; Lu and Clark, 2012; Richardson et al., 2011; Wei et al., 2012). Moreover, the concordance between eQTL direction and logic of miRNA regulation had not been systematically investigated.

To fill in the existing gaps, the next part of the thesis was aimed at identifying and prioritizing SNPs that could influence the expression of the host gene by creating or disrupting the functional miRNA binding site. To this end, a systematic study was performed in which in silico miRNA target site predictions, available *cis*-eQTLs, small RNA sequencing results, Argonaute crosslinking immunoprecipitation (AGO-CLIP) results, and GWAS datasets were integrated.

3.3.1. Identification of 3'-UTR SNPs in *cis*-eQTL loci

To identify SNPs that could influence miRNA binding, it was assumed that the disruption or creation of an miRNA binding site would be reflected at the expression level of the gene harboring the MRE (i.e., *cis*-eQTL effect). This assumption is supported by recent studies postulating that the miRNA targeting effect is reflected in the expression level of the target (Guo et al., 2010).

The first stage of the analysis utilized the largest and most robust set of *cis*-eQTLs available at the time (Westra et al., 2013). In that study, ~5,500 whole blood samples from seven independent cohorts were analyzed by *cis*- and *trans*-eQTL mapping and subsequently integrated in the large-scale meta-analysis. The current study used all 664,097 significant (FDR < 0.05) *cis*-eQTLs (Fig. 13A). In this dataset, a *cis*-eQTL was defined as an association between an eSNP (expression affecting SNP) and nearby (<250 kB) Illumina expression probe. As many genes are detected by several probes from expression arrays, the true biological regulation mechanism is often reflected by several probe-level eQTLs. Perfect proxies ($R^2 = 1$; 1000G pilot 1 data for European population) were calculated for all *cis*-eQTL eSNPs, resulting in the expanded set of 1,198,115 SNP-probe combinations. In animals, most of the highly efficient miRNA binding sites are located in 3' UTRs. Therefore, this study sought to identify *cis*-eQTL SNPs that may influence the expression of the affected gene

in a 3' UTR-mediated manner. Accordingly, *cis*-eQTL SNPs and their proxies were mapped to the 3' UTRs of corresponding *cis*-regulated genes, resulting in 9,334 SNP-probe combinations (0.8% from all *cis*-eQTL effects, Fig. 13B).

About 13% of *cis*-affected transcripts were detected by more than one probe. Among those, ~15% showed a different effect direction from probes detecting the same transcripts. This result may indicate possible alternative splicing or polyadenylation (Zhernakova et al., 2013).

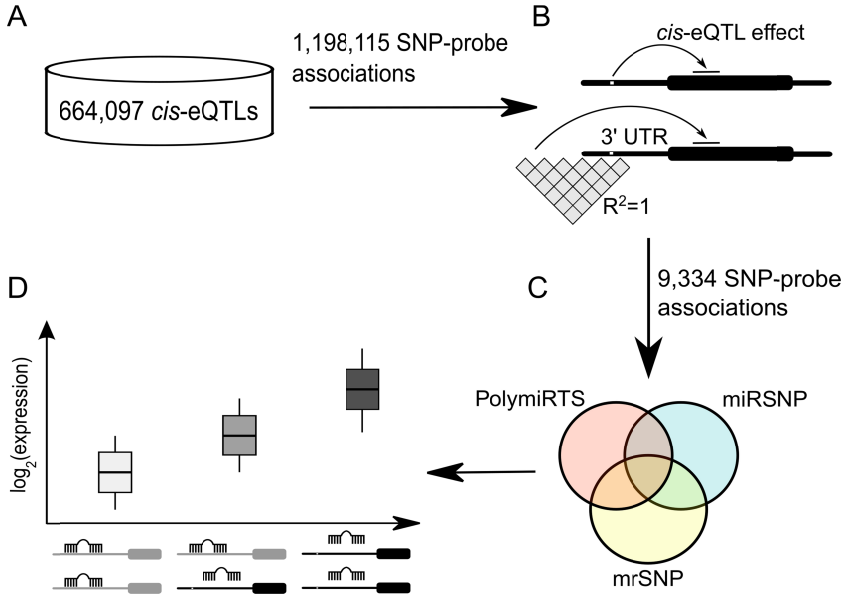


Figure 13. Analysis strategy for identification of miRNA-mediated *cis*-eQTL effects. (A) *cis*-eQTL SNPs and their perfect proxies ($R^2 = 1$, 1000G CEU population) were (B) mapped to the 3' UTRs of *cis*-affected ENSEMBL transcripts. (C) SNPs located within the in silico miRNA binding sites were identified from public databases (PolymiRTS, miRSNP, and webtool mrSNP), each utilizing a different target prediction algorithm. (D) Direction of allelic trends was assessed for concordance with the logic of miRNA-mediated regulation.

3.3.2. Identification of putative MRE-SNPs

To identify SNPs that may influence the binding of miRNAs, all UTR-eQTL SNPs were intersected with three corresponding databases, each utilizing a different target prediction algorithm to predict the effects of the SNP on the MRE (Fig. 13C). To find the strongest effects, only SNPs that were predicted to disrupt, create, or replace the binding site of the corresponding miRNA were included. SNPs predicted to change the strength of miRNA binding were omitted. All target predictions were performed without conservation filters, as several nonconserved MREs have been shown to be functional (Farh et al.,

2005) and using a hard-defined conservation filter may lead to the exclusion of functional target sites (Betel et al., 2010). It is also arbitrary to define in which organisms the MRE has to be conserved to be high confidence.

In the first stage of analysis, the union of all MREs predicted by three different in silico methods was used. About 4% of all *cis*-eQTLs identified from whole blood had an expression associated SNP (eSNP) or its perfect proxy in the in silico-predicted MRE. Those MREs were predicted in genes detected by ~35% of all *cis*-affected probes. Almost all miRNAs present in miRBase v20 had at least one predicted MRE in the set of *cis*-regulated genes (2,753/2,758 miRNAs).

Overall, 27,126 potential miR-SNP-probe associations were identified. However, due to the limitations of prediction algorithms and other factors, those associations likely involve many false positives. Therefore, additional filters were applied, to identify a more confident "filtered set" of putative miRNA-associated *cis*-eQTLs. To add stringency to the in silico target predictions, only MREs predicted by all three prediction algorithms were included. As one of the databases utilizes TargetScan, only those predictions with perfect or near-perfect complementarity in the canonical seed region were included.

To find miRNA-target associations potentially relevant in blood, comprehensive searches were made in GEO for small RNA-seq datasets for whole blood, peripheral blood mononuclear cells, and leukocytes. These searches resulted in 11 datasets, which were reanalyzed by the dedicated tool sRNAbench v0.9. There was substantial heterogeneity among miRNA profiles of the datasets. However, a "consensus" blood-expressed profile of 123 miRNAs could be obtained by using an arbitrary cutoff of at least 10 miRNA reads in at least half of the datasets. Finally, some proportion of miRNAs in the miRBase are falsely annotated (Hansen et al., 2011). Therefore, only the high-confidence set of miRBase v20 miRNAs was used. This set was based on stricter criteria of at least 10 reads aligned to both the 3' and 5' arms of the precursor sequence, as reported by the sequencing studies.

After those criteria were applied, 323 "filtered" miR-SNP-probe combinations were identified. MRE-SNPs were further classified as exclusively MRE-disrupting (minor allele disrupted MRE for all predicted miRNAs), MRE-creating (minor allele created MRE for all predicted miRNAs), or ambiguous (minor allele replaced one set of MREs with another). There was no higher proportion of exclusively MRE-disrupting or MRE-creating MRE-SNPs in the unfiltered and filtered sets of associations (Chi-square test, $P > 0.05$).

3.3.3. Concordance with the logic of miRNA-mediated regulation

Based on the assumption that functional MRE results in the downregulation of the target mRNA, it was hypothesized that disruption of the miRNA binding site by the MRE-SNP would result in upregulation of the target mRNA. Each miR-SNP-probe association was classified as concordant (C-type) or unconcordant (U-type), based on the concordance of the eQTL allelic direction with the logic of miRNA-

meditated regulation (Fig. 13D). *Cis*-eQTLs that were exclusively C-type (allelic direction consistent with the logic of miRNA-mediated regulation for all targeting miRNAs) or exclusively U-type were identified.

The proportion of C-type associations was close to 50% in both unfiltered (49.4%) and filtered miR-SNP-probes (50.6%), indicating no overrepresentation of associations in line with miRNA-mediated regulation. A larger effect size (Z-score) was not observed for C-type associations ($P > 0.05$; Wilcoxon-Mann-Whitney U-test in unfiltered and filtered sets). C-type associations were not overrepresented in the exclusively MRE-creating or exclusively MRE-breaking MRE-SNPs (Chi-squared test $P > 0.05$ for unfiltered and filtered sets). Therefore, there is no indication that MRE-disrupting or MRE-creating SNPs had more pronounced functional consequence on target expression. Similarly, C-type associations were not enriched among conserved compared to unconserved MREs (Chi-square test, $P > 0.05$).

The study next investigated whether C-type relationships are associated with functionally more effective MREs. Several measures of MRE effectiveness were used: free energy (stability of the mRNA-miRNA duplex), miRanda alignment score (alignment between mRNA and miRNA), miRSVR, and context+ scores (two scores constructed from miRNA transfection data by machine learning algorithms). To assess the overall effect of miRNA binding for each miRNA-associated *cis*-eQTL, the average was calculated for each measure, and a comparison was made between exclusively C-type and exclusively U-type *cis*-eQTLs. Higher average effectiveness was not found for C-type addociations in the unfiltered set (Wilcoxon–Mann–Whitney U-test, $P > 0.05$). In the filtered set the mean miRanda score for C-type associations was only marginally lower (Wilcoxon–Mann–Whitney U-test, $P = 0.049$). However, this association contradicts the logic of miRNA-associated regulation. On average, more effective miRNA binding would be expected in the case of C-type associations.

These ambiguous results may indicate that other 3’ UTR-related mechanisms exist that are relevant in the formation of *cis*-eQTL effects. For example, this study did not analyze SNP effects on the binding of other RBPs, on mRNA folding (which may influence accessibility of the MRE for RISC), or on alternative polyadenylation. The two latter mechanisms may influence many regulatory sites and, therefore, exert a large effect. The effect of SNPs on the binding of RBPs is difficult to predict by in silico methods because the binding rules (and, therefore, possible effect of SNPs) of many RBPs are not as well established as the effects of canonical MREs of miRNAs. There is also a complex interplay between the binding of non-AGO RBPs, such as HuR and PUF, and the binding of miRNAs, adding an additional layer of complexity to the gene regulation via the 3’ UTR (Bhattacharyya et al., 2006; Kedde et al., 2010; Kundu et al., 2012). However, a recent study investigated the effects of 3’-UTR SNPs on MREs, mRNA folding, and alternative polyadenylation, using several bioinformatics methods. The results suggested that majority of 3’ UTR SNPs influence the binding of miRNAs, rather than mRNA folding or polyadenylation (Arnold et al., 2012).

3.3.4. Prioritization of putative miRNA-associated *cis*-eQTLs

To prioritize putative miRNA-associated *cis*-eQTLs, the filtered associations were sorted on the basis of the TargetScan context+ score. Several C-type associations were observed among the top prioritized associations (Table 1 of Ref III), including rs10187 in *ISCU* (MRE of miR-210-3p), rs7676 in *C21orf33* (MRE of miR-423-5p), and rs4476230 in *MMD* (MRE of miR-210-3p). However, there were also U-type associations with strong MRE. For example, the second miR-SNP-probe association in the prioritized list was rs11932 in *KPNA1* (MRE of miR-210-3p).

To integrate more supporting evidence in the prioritization of MRE-SNPs, the overlap with the public AGO-CLIP data and experimentally validated MREs were used as supporting evidence. Although ~25% of MRE-SNPs from the unfiltered set of associations were covered by AGO-CLIP reads, there was no overrepresentation of AGO-CLIP-supported *cis*-eQTL effects among the C-type miR-SNP-probe associations (Chi-square test, $P > 0.05$).

Next, three databases of experimentally supported miRNA targets (TarBase, miRTarbase, and miRecords) were used. Whereas 0.85% of miR-SNPs from the unfiltered set of associations affected validated target sites, the number of miR-SNPs in the validated target genes was more than seven times higher among the filtered set of associations (6.5%). For the filtered set of associations overlapping the experimentally supported MREs, the strongest conserved MRE (context+ score -0.554) was associated with rs10187, disrupting the binding site of miR-210-3p in *ISCU* (Fig.14A-C). This binding was supported by AGO-CLIP data and experimental results (Chan et al., 2009; Fasanaro et al., 2009; Helwak et al., 2013; Lee et al., 2011a). *ISCU* functions in mitochondrial iron sulfur cluster assembly, being associated with myalgia in exercise intolerance (Mochel et al., 2008) and decreased cancer survival (Favaro et al., 2010). The miRNA miR-210-3p is a "master hypoxamir" that is regulated by hypoxia-induced factor (reviewed in; (Chan et al., 2012)) and is frequently upregulated in different cancer types, including NSCLC (Ref I, II).

Another prioritized C-type association involved rs4245739, which creates a strong MRE (context+ score -0.32) for miR-191-5p in *MDM4* (Fig. 14D-F). Although this MRE is not directly covered by AGO-CLIP reads, the reads were mapped to the immediate vicinity. *MDM4* encodes an inhibitor of the tumor suppressor p53 and is upregulated in several tumors (Bartel et al., 2005; Han et al., 2007). The minor allele of rs4245739, carried by ~20% of the European population, is associated with a protective effect for several cancers (Eeles et al., 2013; Garcia-Closas et al., 2013; Purrington et al., 2014; Wynendaele et al., 2010). Interestingly, the MRE-creating effect of the SNP to the binding of miR-191-5p and subsequent upregulation of *MDM4* on the transcript and protein levels have been experimentally supported in ovarian cancer cell lines (Wynendaele et al., 2010). This finding represents an example where this analysis strategy independently identified experimentally supported MRE-SNP, integrating information from multiple modalities.

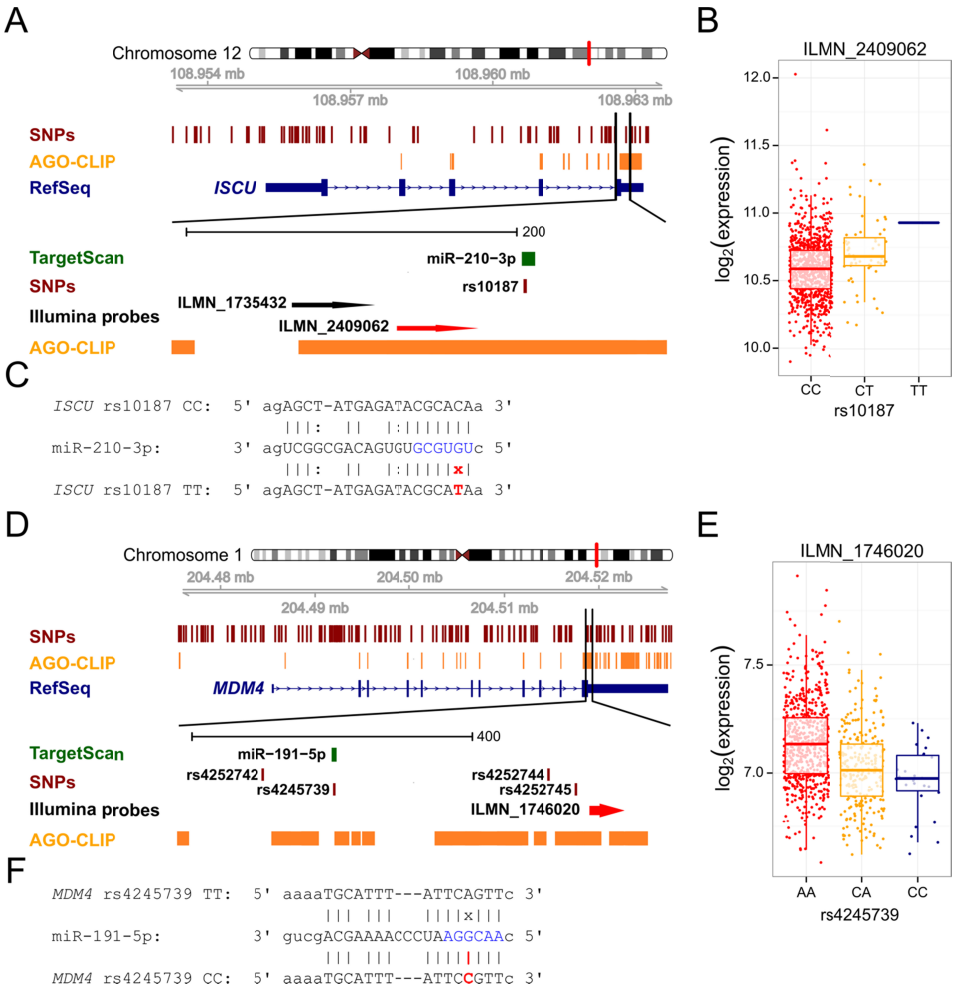


Figure 14. MRE-SNPs in *ISCU* and *MDM4* genes. (a), (d) Schematic depiction of each gene locus. Tracks denote genomic region, SNP positions, AGO binding sites from starBase, most prevalent RefSeq transcript isoform, miRNA target sites from TargetScan v6.2, and Illumina detection probes. Region with the *cis*-eQTL MRE-SNP and affected probe (red) is expanded. (b), (e) Allelic expression pattern in EGCUT sample set. Log-transformed and quantile-normalized expression values are visualized as boxplots. Individual expression values are depicted as points. On the boxplots, the line indicates median, box defines 25–75% quartiles, and whiskers extend the data to 1.5× the interquartile range. (c), (f) MRE-SNP effects on miRNA binding. Vertical lines indicate canonical pairing; colons depict G:U wobble; and "x" denotes a mismatch. Seed region is shown in blue. Minor allele of the MRE-SNP and its effect on miRNA binding are shown in red.

Another SNP, rs2239680, disrupts the MRE of miR-335-5p in the *BIRC5* oncogene. This effect is experimentally supported by results in cell lines (Zu et al., 2013). However, the eQTL direction in blood does not agree with the logic of miRNA-mediated regulation, suggesting that this eQTL may be miRNA-driven only in specific tissue types. The observation that only one of the two experimentally supported MRE-SNP associations was C-type could be caused by the properties of the associated miRNAs. For example, miR-191-5p, which has a C-type association with *MDM4*, is highly expressed in blood (detectable in nine of the 11 publicly available blood samples and representing ~1% of the detected miRNome). The miRNA miR-335-5p was detectable in 6 of the 11 blood samples and represented ~0.1% of the miRNome. At the same time, the in silico-predicted targetome of miR-335-5p was much larger than the one of miR-191-5p (3,046 target genes vs. 568 target genes, respectively; TargetScan v6.2), suggesting that the effect of miR-335-5p may be diluted in this specific tissue type.

3.3.5. Presence of miRNA-associated blood *cis*-eQTLs in other tissues

Publicly available data from the GTEx v4 data portal were used to determine whether the top miRNA-associated *cis*-eQTLs are also detectable in other tissue types, and whether the direction is in agreement with blood eQTLs. Although the sample sizes for other tissues are currently much smaller and the power for proper replication is limited, several associations showed the same allelic direction and nominally significant *cis*-eQTL effects in multiple tissues (Fig. 15). C-type associations showing nominal significance in the largest number of other tissues involved rs1136808 in *CTSS* (MRE of miR-423-5p) and rs2287067 in *VWDE* (MRE of miR-590-5p). Although rs10187 in *ISCU* was not available in the GTEx database, the C-type association involving rs4245739 in *MDM4* was concordant and detectable in several tissues, including the GTEx dataset for blood. This result suggests that those potentially miRNA-related *cis*-eQTLs may have widespread effects in multiple tissue types.

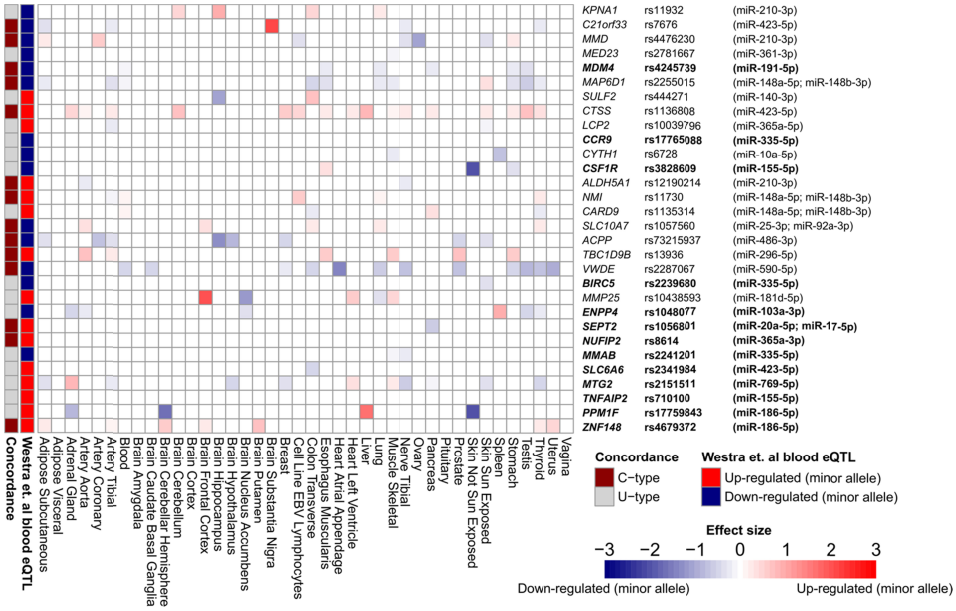


Figure 15. Overlap between top filtered miR-SNP-probe associations and *cis*-eQTLs from the GTEx Consortium. Color of the cell depicts effect size and direction (beta-value). Annotation bars on the left depict association concordance with the logic of miRNA-mediated regulation (C- or U-type) and the effect direction in the study by Westra et al. Only effect sizes for nominally significant tissue eQTLs are shown (uncorrected $P < 0.05$). Associations in bold contain SNPs in experimentally supported miRNA binding sites. Shown are the top 30 filtered miR-SNP-probe associations and associations with validated miRNA binding sites.

3.3.6. Complex trait-associated SNPs in MREs

Sets of unfiltered and filtered MRE-SNPs were intersected with the NHGRI Catalog of Published Genome-Wide Association Studies, to determine whether the identified sets of MRE-associated *cis*-eQTLs can be associated with known genetic risk factors. Among the 5,994 potential MRE-SNPs, 208 (3.5%) overlapped with the 154 GWAS SNPs or their perfect proxies ($R^2 = 1$, 1000G pilot 1, CEU). For the filtered set of miR-SNP-probe associations, 10 MRE-SNPs (including four C-type MRE-SNPs) overlapped with GWAS loci, associated with 12 traits.

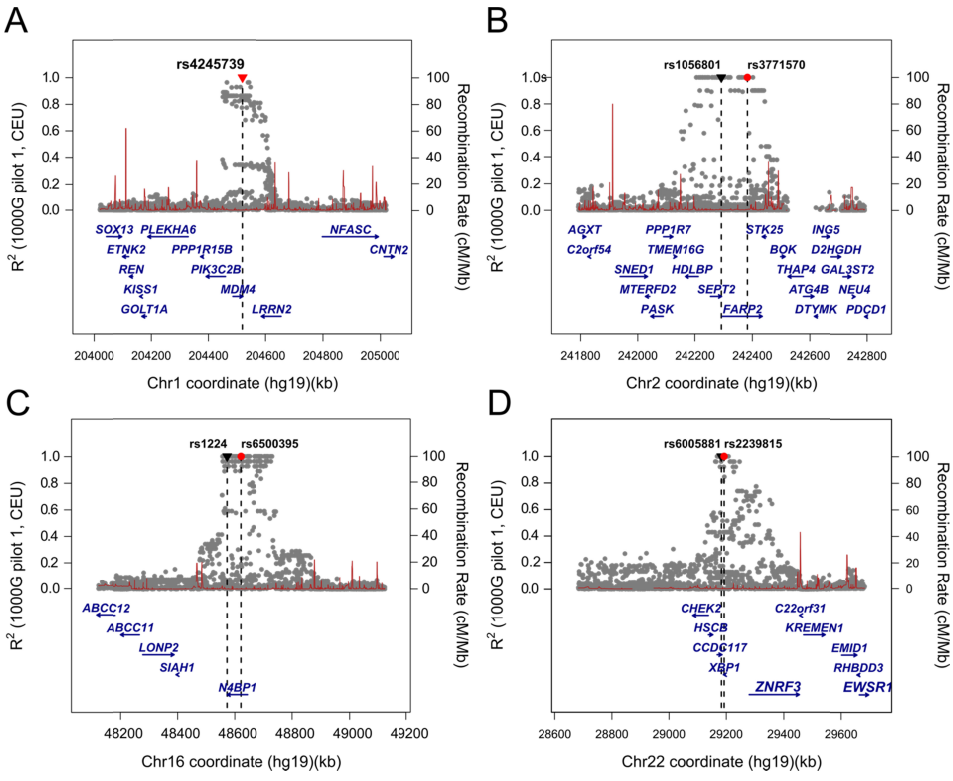


Figure 16. Regional linkage disequilibrium (LD) plots of four C-type MRE-SNPs overlapping with the GWAS Catalogue.

One of the identified MRE-SNPs, rs4245739 in the 3' UTR of *MDM4* (Fig. 16A), is a GWAS SNP associated with decreased risk for several cancers (Eeles et al., 2013; Garcia-Closas et al., 2013; Purrington et al., 2014; Wynendaele et al., 2010). As discussed in Chapter 3.3.4, the MRE-SNP effect of this SNP on miRNA binding has been experimentally verified in cell lines (Wynendaele et al., 2010). A second GWAS SNP, rs6500395, is located in the first intron of *N4BP1* (Fig. 16C) and is associated with the response of rheumatoid arthritis patients to tocilizumab (Wang et al., 2013). However, this MRE-SNP has a perfect proxy (rs1224) in the AGO-CLIP-supported C-type MRE for miR-330-3p.

Two SNPs had GWAS SNPs and perfectly correlating MRE-SNPs in two different genes. The first SNP, rs3771570, is associated with aggressive prostate cancer (Eeles et al., 2013) and is located in the intronic region of *FARP2*. However, its perfect proxy, rs1056801, is positioned in the 3' UTR of the neighboring gene, *SEPT2* (Fig. 16B). The proxy disrupts the binding sites of two members of the oncogenic miR-17-92 family (miR-20a-5p and miR-17-5p), and its host gene is dysregulated in several cancer types (Liu et al., 2010). As *SEPT2* was the only gene with a significant *cis*-eQTL effect from the

corresponding LD-block (R^2 = 1, 1000G pilot 1, CEU), the binding of miR-17-92 family members may be linked to dysregulation of *SEPT2* expression and play a potential role in prostate cancer risk. The second SNP, rs2239815, is associated with esophageal SCC (Wu et al., 2012). Although the SNP itself is positioned in *XBP1*, another perfectly correlating MRE-SNP is in the 3' UTR of *CCDC117* (Fig. 16D). Based on the functional annotations, this locus harbors two potentially cancer-associated genes, *XBP1* and *CHEK2*. Although all three genes are influenced by the eQTLs of this LD-block, the largest effect is for *XBP1* (FDR < 0.01, Z = 23) (Westra et al., 2013), making it unlikely that miRNA-mediated regulation of *CCDC117* plays a role in esophageal SCC.

CONCLUSIONS

The aim of the work presented in this thesis was to untangle the role of miRNAs in the pathogenesis of lung cancer, and to connect genetic variations with miRNA regulatory networks. Using data from a cohort of Estonian NSCLC patients, numerous dysregulated miRNAs were identified, including several known tumorigenic miRNAs and others that have not previously been connected with the pathogenesis of NSCLC. Bioinformatics analyses suggested that the differentially expressed miRNAs are involved in several pathways associated with cell proliferation, differentiation, and motility. The miRNA miR-374a-5p was shown to correlate with the postoperative survival of Estonian NSCLC patients.

These results were systematically combined with information from previous studies by performing a meta-analysis of aberrantly expressed miRNAs in lung cancer. Although a substantial heterogeneity was observed between the results of different studies, a robust set of seven up- and eight downregulated miRNAs was identified, which may serve as potential drug targets or biomarkers in future studies.

To investigate the impact of genetic variants on miRNA regulation, the most comprehensive set of blood *cis*-eQTLs available to date was integrated with several publicly available data sources. For the first time, the concordance between eQTL allelic directions and the logic of miRNA-mediated regulation was analyzed. There was no evident overrepresentation of eQTLs in line with the logic of miRNA-mediated regulation, suggesting that additional mechanisms may play major roles in the formation of eQTL effects in blood. However, after prioritizing the putative miRNA-associated *cis*-eQTLs, several potentially miRNA-driven eQTLs were identified, and one SNP with experimentally supported effects on miRNA binding was independently confirmed.

SUMMARY IN ESTONIAN

MikroRNAde roll haiguste kujunemisel: aberratsioonid kopsuvähis ning seosed genoomse varieeruvusega

Viimase paari kümnendi geneetilised uuringud on leidnud, et suur osa inimese genoomilt transkribeeritavast RNAst valke ei kodeeri. Selline mittekodeeriv RNA omab eelkõige regulatoorset rolli, mõjutades kuidas valke kodeerivad geenid oma funktsioone täidavad. Üks enim uuritud klass mittekodeerivaid RNAsid on mikroRNAd, väikesed ~20 nukleotiidi pikkused RNA molekulid, mis reguleerivad geenide ekspressiooni, seondudes komplementaarsusepõhiselt sihtmärgiks olevate transkriptidega. Iga mikroRNA võib reguleerida sadu sihtmärkgeene, mistõttu on nende molekulide roll bioloogilistes protsessides väga tähtis.

Kuna mikroRNAd mängivad olulist rolli ka mitmete patoloogiliste protsesside tekkes ja arengus, on neil ka suur potentsiaal võimalike ravimisihtmärkidena. Eriti palju funktsionaalseid seoseid on leitud miRNAde ja erinevate kasvajate vahel. Esimene mikroRNA-põhine vähiravim, MRX34, on 2016 aasta seisuga kliiniliste uuringute faasis.

MikroRNAde ekspressioonimuster on vähikudedes muutunud, peegeldades kasvaja molekulaarseid ja histoloogilisi omadusi. Samuti on mikroRNAd detekteeritavad erinevatest kehavedelikest. Nende omaduste tõttu on need molekulid väga huvipakkuvad diagnostiliste ja prognostiliste biomarkeritena mis võimaldaks kasvajate varasemat avastamist ning optimaalsete raviotsuste tegemist.

Käesoleva doktoritöö eesmärk oli uurida mikroRNAde rolli mitteväikerakulise kopsuvähi patoloogias, identifitseerida uusi potentsiaalseid prognostilisi ja diagnostilisi biomarkereid ning leida seoseid mikroRNAde toimemehhanismide ning inimese genoomi normaalse geneetilise varieeruvuse vahel. Uuritud vähitüüp on Eestis sageduselt teine ning peamine vähisurmade põhjustaja. Antud doktoritöös kasutati koostöös Tartu Ülikooli Kopsukliinikuga kogutud algstaadiumi kopsuvähi proove ning viidi läbi mikroRNAde profileerimine. Analüüside tulemusena leiti 72 mikroRNAd mis olid oluliselt üles- või alla reguleeritud võrrelduna vähivaba kopsukoega ja võivad seetõttu huvi pakkuda diagnostiliste biomarkerite või ravimisihtmärkidena. Leitud mikroRNAde ennustuslikud sihtmärgid olid seotud mitmete vähiga seotud bioloogiliste protsessidega, toetades nende potentsiaalset rolli vähi arengus. Lisaks identifitseeriti üks mikroRNA mis näitas uuritud kopsuvähi kohordis assotsiatsiooni operatsioonijärgse elulemusega.

Vaatamata sellele et teise generatsiooni sekveneerimismeetodeid kasutatakse viimastel aastatel üha laialdasemalt, on suur enamus mikroRNAde profileerimise uuringuid endiselt läbi viidud odavamate ekspressioonikiibi tehnoloogiate abil. Samas on avaldatud mikroRNAde ekspressioonianalüüside tulemused suhteliselt varieeruvad ja vahel isegi vastukäivad, muutes bioloogiliste järelduste tegemise ning kõige huvipakkuvamate mikroRNAde leidmise keerukaks ülesandeks. Võrdlemaks käesolevas uuringus leitud mikroRNAde

vähispetsiifilist profiili teiste sarnaste töödega, viidi läbi süstemaatiline infootsing ning meta-analüüs. Kuna erinevad mikroRNAde ekspressiooni tuvastamise meetodid pole omavahel otseselt võrreldavad, kasutati meta-analüüsiks spetsiaalset järjestatud geenilistide analüüsimise meetodit. Kahekümne uuringusse kaasatud andmestiku meta-analüüsi tulemusena leiti seitse mikroRNAd mis olid kopsuvähis üles- ja kaheksa mis olid alla reguleeritud. See profiil on märksa usaldusväärsem kui iga üksiku uuringu poolt leitu ning sobib lähtepunktiks järgnevatele funktsionaalsetele uuringutele.

Ülegenoomsed assotsiatsiooniuuringud on identifitseerinud tuhandeid inimpopulatsioonis leiduvaid ühenukleotiidseid polümorfisme (*single nucleotide polymorphism*- SNP) mis on assotsieerunud erinevate haiguste ja fenotüübiliste tunnustega. Paraku on bioloogiline mehhanism, kuidas need variandid oma mõju fenotüübile avaldavad, enamikul juhtudel siiani teadmata. Valdav osa seda tüüpi variante asuvad genoomi mittekodeerivates alades, mõjutades tõenäoliselt pigem geenide ekspressioonitasemeid kui geeni produktiks oleva valgu struktuuri. Kuna mikroRNAde seondumisalades asuvad SNPd võivad mõjutada vastava mikroRNA seondumist ja regulatoorset efektiivsust, siis on see ka üks võimalik mehhanism geeni ekspressiooni ja fenotüübil mõjutamiseks.

Käesolevas uuringus seoti mikroRNAde regulatoorne potentsiaal inimese normaalse geneetilise varieeruvusega, kaardistades selleks teadaolevad geeniekspressiooni mõjutavad SNPd mikroRNAde ennustuslikesse seondumisaladesse. Esmakordselt testiti süsteemselt, kas geeniekspressiooni muutuse põhjuseks võib olla mikroRNAde toime. Selleks eeldati, et funktsionaalne mikroRNA seondumine surub sihtmärkgeeni ekspressiooni alla ning iga assotsiatsiooni ekspressiooni muutust võrreldi mainitud toimemehhanismi loogikaga. Lisaks integreeriti mitmeid avalikult kättesaadavaid mikroRNAde ekspressioonitasemete ja mikroRNAdega kompleksis olevate valkude immunopretsipitatsiooni andmestikke, suurendamaks identifitseeritud assotsiatsioonide usaldusväärsust.

Uuringu tulemusel ei leitud mikroRNAde toimemehhanismi loogikaga kooskõlas olevate assotsiatsioonide üle-esindatust, mistõttu on põhjust arvata, et paljudel juhtudel mõjutavad potentsiaalsetes mikroRNAde seondumisalades asuvad genoomsed variandid geeniekspressiooni teiste mehhanismide kaudu. Sellisteks mehhanismideks võivad olla näiteks alternatiivne polüadenülatsioon, alternatiivne mRNA voltumine või erinevate RNAga seonduvate valkude seondumine. Samas leiti antud uuringus mitmeid genoomseid variante, mille osalust geeniekspressiooni regulatsioonis just nimelt mikroRNAde kaudu toetasid mitmed andmed. Identifitseeriti ka neli juhtu kus mikroRNAde regulatsiooni loogikaga kooskõlas olev genoomne variant asus haigusriskiga seotud lookuses.

Kokkuvõtteks võib öelda et antud doktoritöö identifitseeris mitmeid mikroRNAsid mis näitasid seost kopsuvähi patogeneesiga, aitas süstematiseerida varasemate uuringute tulemusi ning sidus mikroRNAde regulatoorse potentsiaali normaalse geneetilise varieeruvusega.

REFERENCES

Abelson, J.F., Kwan, K.Y., Roak, B.J.O., Baek, D.Y., Stillman, A.A., Morgan, T.M., and State, M.W. (2005). Sequence Variants in SLITRK1 Are Associated with Tourette ' s Syndrome. Science *2126*, 317–320.

Agarwal, V., Bell, G.W., Nam, J.-W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. Elife *4*, e05005.

Ali, S., Saleh, H., Sethi, S., Sarkar, F.H., and Philip, P. a (2012). MicroRNA profiling of diagnostic needle aspirates from patients with pancreatic cancer. Br. J. Cancer *107*, 1354–1360.

Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T., and Jewell, D. (2003). MicroRNAs and other tiny endogenous RNAs in C. Elegans. Curr. Biol. *13*, 1317–1323.

Arnold, M., Ellwanger, D.C., Hartsperger, M.L., Pfeufer, A., and Stümpflen, V. (2012). Cis-Acting Polymorphisms Affect Complex Traits through Modifications of MicroRNA Regulation Pathways. PLoS One *7*, e36694.

Arroyo, J.D., Chevillet, J.R., Kroh, E.M., Ruf, I.K., Pritchard, C.C., Gibson, D.F., Mitchell, P.S., Bennett, C.F., Pogosova-Agadjanyan, E.L., Stirewalt, D.L., et al. (2011). Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. Proc. Natl. Acad. Sci. *108*, 5003–5008.

Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P., and Blelloch, R. (2008). Mouse ES cells express endogenous shRNAs, siRNAs, and other microprocessor-independent, dicer-dependent small RNAs. Genes Dev. *22*, 2773–2785.

Baek, D., Villén, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. Nature *455*, 64–71.

Bagga, S., Bracht, J., Hunter, S., Massirer, K., Holtz, J., Eachus, R., and Pasquinelli, A.E. (2005). Regulation by let-7 and lin-4 miRNAs Results in Target mRNA Degradation. Cell *122*, 553–563.

Bai, Z.-Q., Lei, X.-F., Wang, L.-D., and Gao, S.-J. (2008). Identification and Function of MicroRNAs Encoded by Herpesviruses. Virol. Sin. *23*, 459–472.

Bao, L., Zhou, M., Wu, L., Lu, L., Goldowitz, D., Williams, R.W., and Cui, Y. (2007). PolymiRTS Database: linking polymorphisms in microRNA target sites with complex traits. Nucleic Acids Res. *35*, D51–D54.

Barenboim, M., Zoltick, B.J., Guo, Y., and Weinberger, D.R. (2010). MicroSNiPer: a web tool for prediction of SNP effects on putative microRNA targets. Hum. Mutat. *31*, 1223–1232.

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

Bartel, F., Schulz, J., Böhnke, A., Blümke, K., Kappler, M., Bache, M., Schmidt, H., Würl, P., Taubert, H., and Hauptmann, S. (2005). Significance of HDMX-S (or MDM4) mRNA splice variant overexpression and HDMX gene amplification on primary soft tissue sarcoma prognosis. Int. J. Cancer *117*, 469–475.

Baskerville, S., and Bartel, D.P. (2005). Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA *11*, 241–247.

Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P., and Izaurralde, E. (2006). mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. Genes Dev. *20*, 1885–1898.

Berezikov, E., Chung, W.-J., Willis, J., Cuppen, E., and Lai, E.C. (2007). Mammalian mirtron genes. Mol. Cell *28*, 328–336.

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

Betel, D., Koppal, A., Agius, P., Sander, C., and Leslie, C. (2010). Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol. *11*, R90.

Bhattacharya, A., Ziebarth, J.D., and Cui, Y. (2014). PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. Nucleic Acids Res. *42*, D86–D91.

Bhattacharyya, S.N., Habermacher, R., Martine, U., Closs, E.I., and Filipowicz, W. (2006). Relief of microRNA-Mediated Translational Repression in Human Cells Subjected to Stress. Cell *125*, 1111–1124.

Boeri, M., Verri, C., Conte, D., Roz, L., Modena, P., Facchinetti, F., Calabrò, E., Croce, C.M., Pastorino, U., and Sozzi, G. (2011). MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. Proc. Natl. Acad. Sci. *108*, 3713–3718.

Borel, C., Deutsch, S., Letourneau, A., Migliavacca, E., Montgomery, S.B., Dimas, A.S., Vejnar, C.E., Attar, H., Gagnebin, M., Gehrig, C., et al. (2011). Identification of cis- and trans-regulatory variation modulating microRNA expression levels in human fibroblasts. Genome Res. *21*, 68–73.

Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005). Principles of MicroRNA–Target Recognition. PLoS Biol. *3*, e85.

Bulik-Sullivan, B., Selitsky, S., and Sethupathy, P. (2013). Prioritization of Genetic Variants in the microRNA Regulome as Functional Candidates in Genome-Wide Association Studies. Hum. Mutat. *34*, 1049–1056.

Burroughs, A.M., Ando, Y., De Hoon, M.J.L., Tomaru, Y., Nishibu, T., Ukekawa, R., Funakoshi, T., Kurokawa, T., Suzuki, H., Hayashizaki, Y., et al. (2010). A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. Genome Res. *20*, 1398–1410.

Cai, L., and Cai, X. (2014). Up-regulation of miR-9 expression predicate advanced clinicopathological features and poor prognosis in patients with hepatocellular carcinoma. Diagn. Pathol. *9*, 1000.

Cai, J., Guan, H., Fang, L., Yang, Y., Zhu, X., Yuan, J., Wu, J., and Li, M. (2013). MicroRNA-374a activates Wnt/β-catenin signaling to promote breast cancer metastasis. J. Clin. Invest. *123*, 566–579.

Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., et al. (2002). Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc. Natl. Acad. Sci. *99*, 15524–15529.

Calin, G.A., Sevignani, C., Dumitru, C.D., Hyslop, T., Noch, E., Yendamuri, S., Shimizu, M., Rattan, S., Bullrich, F., Negrini, M., et al. (2004). Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc. Natl. Acad. Sci. *101*, 2999–3004.

Calin, G.A., Ferracin, M., Cimmino, A., Di Leva, G., Shimizu, M., Wojcik, S.E., Iorio, M. V, Visone, R., Sever, N.I., Fabbri, M., et al. (2005). A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. N. Engl. J. Med. *353*, 1793–1801.

Campayo, M., Navarro, A., Viñolas, N., Tejero, R., Muñoz, C., Diaz, T., Marrades, R., Cabanas, M.L., Gimferrer, J.M., Gascon, P., et al. (2011). A dual role for KRT81: A miR-SNP associated with recurrence in Non-Small-Cell lung cancer and a novel marker of squamous cell lung carcinoma. PLoS One *6*, 1–9.

Castoldi, M., Schmidt, S., Benes, V., Noerholm, M., Kulozik, A.E., Hentze, M.W., and Muckenthaler, M.U. (2006). A sensitive array for microRNA expression profiling (miChip) based on locked nucleic acids (LNA). RNA *12* , 913–920.

Chan, S.Y., Zhang, Y.-Y., Hemann, C., Mahoney, C.E., Zweier, J.L., and Loscalzo, J. (2009). MicroRNA-210 Controls Mitochondrial Metabolism during Hypoxia by Repressing the Iron-Sulfur Cluster Assembly Proteins ISCU1/2. Cell Metab. *10*, 273–284.

Chan, Y.C., Banerjee, J., Choi, S.Y., and Sen, C.K. (2012). miR-210: The Master Hypoxamir. Microcirculation *19*, 215–223.

Chandra, V., Girijadevi, R., Nair, A.S., Pillai, S.S., and Pillai, R.M. (2010). MTar: a computational microRNA target prediction architecture for human transcriptome. BMC Bioinformatics *11*, S2.

Cheloufi, S., Dos Santos, C.O., Chong, M.M.W., and Hannon, G.J. (2010). A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature *465*, 584–589.

Chen, C., Ridzon, D.A., Broomer, A.J., Zhou, Z., Lee, D.H., Nguyen, J.T., Barbisin, M., Xu, N.L., Mahuvakar, V.R., Andersen, M.R., et al. (2005). Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res. *33*, e179.

Chen, Q., Jiao, D., Yan, L., Wu, Y., Hu, H., Song, J., Yan, J., Wu, L., Xu, L., and Shi, J. (2015). Comprehensive gene and microRNA expression profiling reveals miR-206 inhibits MET in lung cancer metastasis. Mol. BioSyst. *11*, 2290–2302.

Chen, X., Ba, Y., Ma, L., Cai, X., Yin, Y., Wang, K., Guo, J., Zhang, Y., Chen, J., Guo, X., et al. (2008). Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res. *18*, 997–1006.

Chen, Y., Zhu, X., Zhang, X., Liu, B., and Huang, L. (2010). Nanoparticles Modified With Tumor-targeting scFv Deliver siRNA and miRNA for Cancer Therapy. Mol. Ther. *18*, 1650–1656.

Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., and Shiekhattar, R. (2005). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature *436*, 740–744.

Chendrimada, T.P., Finn, K.J., Ji, X., Baillat, D., Gregory, R.I., Liebhaber, S. a, Pasquinelli, A.E., and Shiekhattar, R. (2007). MicroRNA silencing through RISC recruitment of eIF6. Nature *447*, 823–828.

Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature *460*, 479–486.

Chi, S.W., Hannon, G.J., and Darnell, R.B. (2012). An alternative mode of microRNA target recognition. Nat. Struct. Mol. Biol. *19*, 321–327.

Chiang, H.R., Schoenfeld, L.W., Ruby, J.G., Auyeung, V.C., Spies, N., Baek, D., Johnston, W.K., Russ, C., Luo, S., Babiarz, J.E., et al. (2010). Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev. *24*, 992–1009.

Chin, L.J., Ratner, E., Leng, S., Zhai, R., Nallur, S., Babar, I., Muller, R.U., Straka, E., Su, L., Burki, E.A., et al. (2008). A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk. Cancer Res. *68*, 8535–8540.

Cho, W.C., Chow, A.S., and Au, J.S. (2011). MiR-145 inhibits cell proliferation of human lung adenocarcinoma by targeting EGFR and NUDT1. RNA Biol. *8*, 125–131.

Cho, W.C.S., Chow, A.S.C., and Au, J.S.K. (2009). Restoration of tumour suppressor hsa-miR-145 inhibits cancer cell growth in lung adenocarcinoma patients with epidermal growth factor receptor mutation. Eur. J. Cancer *45*, 2197–2206.

Chou, C.-H., Chang, N.-W., Shrestha, S., Hsu, S.-D., Lin, Y.-L., Lee, W.-H., Yang, C.-D., Hong, H.-C., Wei, T.-Y., Tu, S.-J., et al. (2016). miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic Acids Res. *44* , D239–D247.

Cifuentes, D., Xue, H., Taylor, D.W., Patnode, H., Mishima, Y., Cheloufi, S., Ma, E., Mane, S., Hannon, G.J., Lawson, N.D., et al. (2010). A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. Science *328*, 1694–1698.

Clop, A., Marcq, F., Takeda, H., Pirottin, D., Tordoir, X., Bibé, B., Bouix, J., Caiment, F., Elsen, J.-M., Eychenne, F., et al. (2006). A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nat. Genet. *38*, 813–818.

Crawford, M., Batte, K., Yu, L., Wu, X., Nuovo, G.J., Marsh, C.B., Otterson, G.A., and Nana-Sinkam, S.P. (2009). MicroRNA 133B targets pro-survival molecules MCL-1 and BCL2L2 in lung cancer. Biochem. Biophys. Res. Commun. *388*, 483–489.

Dahiya, N., Sherman-Baust, C.A., Wang, T.L., Davidson, B., Shih, L.M., Zhang, Y., Wood, W., Becker, K.G., and Morin, P.J. (2008). MicroRNA expression and identification of putative miRNA targets in ovarian cancer. PLoS One *3*, e2436.

Davoren, P. a, McNeill, R.E., Lowery, A.J., Kerin, M.J., and Miller, N. (2008). Identification of suitable endogenous control genes for microRNA gene expression analysis in human breast cancer. BMC Mol. Biol. *9*, 76.

Deveci, M., Çatalyürek, Ü. V, and Toland, A. (2014). mrSNP: Software to detect SNP effects on microRNA binding. BMC Bioinformatics *15*, 73.

Doench, J.G., and Sharp, P.A. (2004). Specificity of microRNA target selection in translational repression. Genes Dev. *504*, 504–511.

Duan, S., Mi, S., Zhang, W., and Dolan, M.E. (2009). Comprehensive analysis of the impact of SNPs and CNVs on human microRNAs and their regulatory genes. RNA Biol. *6*, 412–425.

Dvinge, H., Git, A., Gräf, S., Salmon-Divon, M., Curtis, C., Sottoriva, A., Zhao, Y., Hirst, M., Armisen, J., Miska, E. a, et al. (2013). The shaping and functional consequences of the microRNA landscape in breast cancer. Nature *497*, 378–382.

Ebert, M.S., and Sharp, P.A. (2012). Roles for MicroRNAs in Conferring Robustness to Biological Processes. Cell *149*, 515–524.

Ebert, M.S., Neilson, J.R., and Sharp, P.A. (2007). MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. Nat. Methods *4*, 721–726.

Eeles, R. a, Olama, A.A. Al, Benlloch, S., Saunders, E.J., Leongamornlert, D.A., Tymrakiewicz, M., Ghoussaini, M., Luccarini, C., Dennis, J., Jugurnauth-Little, S., et al. (2013). Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nat. Genet. *45*, 385–391, 391e1–e2.

Eichhorn, S.W., Guo, H., McGeary, S.E., Rodriguez-Mias, R.A., Shin, C., Baek, D., Hsu, S., Ghoshal, K., Villén, J., and Bartel, D.P. (2014). mRNA Destabilization Is the Dominant Effect of Mammalian MicroRNAs by the Time Substantial Repression Ensues. Mol. Cell *56*, 104–115.

Ender, C., Krek, A., Friedländer, M.R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N., and Meister, G. (2008). A Human snoRNA with MicroRNA-Like Functions. Mol. Cell *32*, 519–528.

Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in Drosophila. Genome Biol. *5*, R1.

Fan, X., and Kurgan, L. (2015). Comprehensive overview and assessment of computational prediction of microRNA targets in animals. Brief. Bioinform. *16*, 780–794.

Fang, Z., and Rajewsky, N. (2011). The Impact of miRNA Target Sites in Coding Sequences and in 3′UTRs. PLoS One *6*, e18067.

Farh, K.K.-H., Grimson, A., Jan, C., Lewis, B.P., Johnston, W.K., Lim, L.P., Burge, C.B., and Bartel, D.P. (2005). The Widespread Impact of Mammalian MicroRNAs on mRNA Repression and Evolution. Science *310*, 1817–1821.

Fasanaro, P., Greco, S., Lorenzi, M., Pescatori, M., Brioschi, M., Kulshreshtha, R., Banfi, C., Stubbs, A., Calin, G.A., Ivan, M., et al. (2009). An Integrated Approach for Experimental Target Identification of Hypoxia-induced miR-210. J. Biol. Chem. *284*, 35134–35143.

Favaro, E., Ramachandran, A., McCormick, R., Gee, H., Blancher, C., Crosby, M., Devlin, C., Blick, C., Buffa, F., Li, J.L., et al. (2010). MicroRNA-210 regulates mitochondrial free radical response to hypoxia and krebs cycle in cancer cells by targeting iron sulfur cluster protein ISCU. PLoS One *5*, e10345.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2015). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int. J. Cancer *136*, E359–E386.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806–811.

Friedman, R.C., Farh, K.K.-H., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. *19*, 92–105.

Gamazon, E.R., Ziliak, D., Im, H.K., LaCroix, B., Park, D.S., Cox, N.J., and Huang, R.S. (2012). Genetic Architecture of MicroRNA Expression: Implications for the Transcriptome and Complex Traits. Am. J. Hum. Genet. *90*, 1046–1063.

Gao, J., Wang, W.-Y., Mao, Y.-W., Gräff, J., Guan, J.-S., Pan, L., Mak, G., Kim, D., Su, S.C., and Tsai, L.-H. (2010a). A novel pathway regulates memory and plasticity via SIRT1 and miR-134. Nature *466*, 1105–1109.

Gao, W., Yu, Y., Cao, H., Shen, H., Li, X., Pan, S., and Shu, Y. (2010b). Deregulated expression of miR-21, miR-143 and miR-181a in non small cell lung cancer is related to clinicopathologic characteristics or patient prognosis. Biomed. Pharmacother. *64*, 399–408.

Gao, W., Shen, H., Liu, L., Xu, J., Xu, J., and Shu, Y. (2011). MiR-21 overexpression in human primary squamous cell lung carcinoma is associated with poor patient prognosis. J. Cancer Res. Clin. Oncol. *137*, 557–566.

Garcia, D.M., Baek, D., Shin, C., Bell, G.W., Grimson, A., and Bartel, D.P. (2011). Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. Nat. Struct. Mol. Biol. *18*, 1139–1146.

Garcia-Closas, M., Couch, F.J., Lindstrom, S., Michailidou, K., Schmidt, M.K., Brook, M.N., Orr, N., Rhie, S.K., Riboli, E., Feigelson, H.S., et al. (2013). Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat. Genet. *45*, 392–398, 398e1–e2.

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science *312*, 75–79.

Gong, J., Tong, Y., Zhang, H.-M., Wang, K., Hu, T., Shan, G., Sun, J., and Guo, A.-Y. (2012). Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. Hum. Mutat. *33*, 254–263.

Gong, J., Liu, C., Liu, W., Wu, Y., Ma, Z., Chen, H., and Guo, A.-Y. (2015). An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools. Database *2015*.

Goya, T., Asamura, H., Yoshimura, H., Kato, H., Shimokata, K., Tsuchiya, R., Sohara, Y., Miya, T., and Miyaoka, E. (2005). Prognosis of 6644 resected non-small cell lung cancers in Japan: A Japanese lung cancer registry study. Lung Cancer *50*, 227–234.

Griffiths-Jones, S. (2004). The microRNA Registry. Nucleic Acids Res. *32*, 109D – 111.

Griffiths-Jones, S. (2006). miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. *34*, D140–D144.

Grimson, A., Farh, K.K.-H., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol. Cell *27*, 91–105.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell *106*, 23–34.

Grosswendt, S., Filipchyk, A., Manzano, M., Klironomos, F., Schilling, M., Herzog, M., Gottwein, E., and Rajewsky, N. (2014). Unambiguous Identification of miRNA: Target Site Interactions by Different Types of Ligation Reactions. Mol. Cell *54*, 1042–1054.

Guduric-Fuchs, J., O'Connor, A., Camp, B., O'Neill, C.L., Medina, R.J., and Simpson, D. a (2012). Selective extracellular vesicle-mediated export of an overlapping set of microRNAs from multiple cell types. BMC Genomics *13*, 357.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840.

Guo, J.U., Agarwal, V., Guo, H., and Bartel, D.P. (2014). Expanded identification and characterization of mammalian circular RNAs. Genome Biol. *15*, 409.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.C., Munschauer, M., et al. (2010). Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. Cell *141*, 129–141.

Hamilton, a J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science *286*, 950–952.

Han, B.W., Hung, J.-H., Weng, Z., Zamore, P.D., and Ameres, S.L. (2011). The 3′-to-5′ Exoribonuclease Nibbler Shapes the 3′ Ends of MicroRNAs Bound to Drosophila Argonaute1. Curr. Biol. *21*, 1878–1887.

Han, J., Lee, Y., Yeom, K.-H., Kim, Y.-K., Jin, H., and Kim, V.N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev. *18*, 3016–3027.

Han, X., Garcia-Manero, G., McDonnell, T.J., Lozano, G., Medeiros, L.J., Xiao, L., Rosner, G., Nguyen, M., Fernandez, M., Valentin-Vega, Y. a, et al. (2007). HDM4

(HDMX) is widely expressed in adult pre-B acute lymphoblastic leukemia and is a potential therapeutic target. Mod. Pathol. *20*, 54–62.

Hanke, M., Hoefig, K., Merz, H., Feller, A.C., Kausch, I., Jocham, D., Warnecke, J.M., and Sczakiel, G. (2010). A robust methodology to study urine microRNA as tumor marker: MicroRNA-126 and microRNA-182 are related to urinary bladder cancer. Urol. Oncol. Semin. Orig. Investig. *28*, 655–661.

Hansen, T.B., Kjems, J., and Bramsen, J.B. (2011). Enhancing miRNA annotation confidence in miRBase by continuous cross dataset analysis. RNA Biol. *8*, 1–6.

Havens, M.A., Reich, A.A., Duelli, D.M., and Hastings, M.L. (2012). Biogenesis of mammalian microRNAs by a non-canonical processing pathway. Nucleic Acids Res. *40*, 4626–4640.

He, H., Jazdzewski, K., Li, W., Liyanarachchi, S., Nagy, R., Volinia, S., Calin, G.A., Liu, C.G., Franssila, K., Suster, S., et al. (2005a). The role of microRNA genes in papillary thyroid carcinoma. Proc. Natl. Acad. Sci. *102*, 19075–19080.

He, L., Thomson, J.M., Hemann, M.T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S.W., Hannon, G.J., et al. (2005b). A microRNA polycistron as a potential human oncogene. Nature *435*, 828–833.

He, W., Feng, L., Xia, D., and Han, N. (2015). MiR-374a promotes the proliferation of human osteosarcoma by downregulating FOXO1 expression. Int. J. Clin. Exp. Med. *8*, 3482–3489.

Heimberg, A.M., Sempere, L.F., Moy, V.N., Donoghue, P.C.J., and Peterson, K.J. (2008). MicroRNAs and the advent of vertebrate morphological complexity. Proc. Natl. Acad. Sci. *105*, 2946–2950.

Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D. (2013). Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding. Cell *153*, 654–665.

Hiard, S., Charlier, C., Coppieters, W., Georges, M., and Baurain, D. (2010). Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. Nucleic Acids Res. *38*, D640–D651.

Hindorff, L. a, Sethupathy, P., Junkins, H. a, Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T. a (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. *106*, 9362–9367.

Höck, J., and Meister, G. (2008). The Argonaute protein family. Genome Biol. *9*, 210.

Houston, K.A., Henley, S.J., Li, J., White, M.C., and Richards, T.B. (2014). Patterns in lung cancer incidence rates and trends by histologic type in the United States, 2004-2009. Lung Cancer *86*, 22–28.

Hsu, S. Da, Tseng, Y.T., Shrestha, S., Lin, Y.L., Khaleel, A., Chou, C.H., Chu, C.F., Huang, H.Y., Lin, C.M., Ho, S.Y., et al. (2014). MiRTarBase update 2014: An information resource for experimentally validated miRNA-target interactions. Nucleic Acids Res. *42*, 78–85.

Huan, T., Rong, J., Liu, C., Zhang, X., Tanriverdi, K., Joehanes, R., Chen, B.H., Murabito, J.M., Yao, C., Courchesne, P., et al. (2015). Genome-wide identification of microRNA expression quantitative trait loci. Nat. Commun. *6*, 6601.

Huang, R.S., Gamazon, E.R., Ziliak, D., Wen, Y., Im, H.K., Zhang, W., Wing, C., Duan, S., Bleibel, W.K., Cox, N.J., et al. (2011). Population differences in microRNA expression and biological implications. RNA Biol. *8*, 692–701.

Huang, X., Le, Q.-T., and Giaccia, A.J. (2010). MiR-210 – micromanager of the hypoxia pathway. Trends Mol. Med. *16*, 230–237.

Humphreys, D.T., Westman, B.J., Martin, D.I.K., and Preiss, T. (2005). MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. Proc. Natl. Acad. Sci. *102*, 16961–16966.

Hutvágner, G., McLachlan, J., Pasquinelli, A.E., Bálint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science *293*, 834–838.

Iorio, M. V, Ferracin, M., Liu, C.G., Veronese, A., Spizzo, R., and Sabbioni, S. (2005). MicroRNA gene expression deregulation in human breast cancer. Cancer Res *65*.

Izzotti, A., Calin, G.A., Steele, V.E., Croce, C.M., and De Flora, S. (2009a). Relationships of microRNA expression in mouse lung with age and exposure to cigarette smoke and light. FASEB J. *23*, 3243–3250.

Izzotti, A., Calin, G.A., Arrigo, P., Steele, V.E., Croce, C.M., and De Flora, S. (2009b). Downregulation of microRNA expression in the lungs of rats exposed to cigarette smoke. FASEB J. *23*, 806–812.

Jang, J.S., Jeon, H.-S., Sun, Z., Aubry, M.C., Tang, H., Park, C.-H., Rakhshan, F., Schultz, D.A., Kolbert, C.P., Lupu, R., et al. (2012). Increased miR-708 Expression in NSCLC and Its Association with Poor Survival in Lung Adenocarcinoma from Never Smokers. Clin. Cancer Res. *18*, 3658–3667.

John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., and Marks, D.S. (2004). Human MicroRNA Targets. PLoS Biol. *2*, e363.

Johnson, D.H., Fehrenbacher, L., Novotny, W.F., Herbst, R.S., Nemunaitis, J.J., Jablons, D.M., Langer, C.J., DeVore, R.F., Gaudreault, J., Damico, L.A., et al. (2004). Randomized phase II trial comparing bevacizumab plus carboplatin and paclitaxel with carboplatin and paclitaxel alone in previously untreated locally advanced or metastatic non-small-cell lung cancer. J. Clin. Oncol. *22*, 2184–2191.

Johnson, S.M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K.L., Brown, D., and Slack, F.J. (2005). RAS is regulated by the let-7 microRNA family. Cell *120*, 635–647.

Junker, A., Krumbholz, M., Eisele, S., Mohan, H., Augstein, F., Bittner, R., Lassmann, H., Wekerle, H., Hohlfeld, R., and Meinl, E. (2009). MicroRNA profiling of multiple sclerosis lesions identifies modulators of the regulatory protein CD47. Brain *132*, 3342–3352.

Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A.G., and Nishikura, K. (2007). Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science *315*, 1137–1140.

Kedde, M., van Kouwenhove, M., Zwart, W., Oude Vrielink, J.A.F., Elkon, R., and Agami, R. (2010). A Pumilio-induced RNA structure switch in p27-3[prime] UTR controls miR-221 and miR-222 accessibility. Nat Cell Biol *12*, 1014–1020.

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. (2007). The role of site accessibility in microRNA target recognition. Nat. Genet. *39*, 1278–1284.

Ketting, R.F., Fischer, S.E.J., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H.A. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev. *15*, 2654–2659.

Khorshid, M., Hausser, J., Zavolan, M., and van Nimwegen, E. (2013). A biophysical miRNA-mRNA interaction model infers canonical and noncanonical targets. Nat. Methods *10*, 253–255.

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209–216.

Kim, D.H., SÃ¦trom, Pã¥., SnÃ¸ve, O., and Rossi, J.J. (2008). MicroRNA-directed transcriptional gene silencing in mammalian cells. Proc. Natl. Acad. Sci. *105*, 16230–16235.

Kiriakidou, M. (2004). A combined computational-experimental approach predicts human microRNA targets. Genes Dev. *18*, 1165–1178.

Kolde, R., Laur, S., Adler, P., and Vilo, J. (2012). Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics *28*, 573–580.

Kota, J., Chivukula, R.R., O'Donnell, K.A., Wentzel, E.A., Montgomery, C.L., Hwang, H.W., Chang, T.C., Vivekanandan, P., Torbenson, M., Clark, K.R., et al. (2009). Therapeutic microRNA Delivery Suppresses Tumorigenesis in a Murine Liver Cancer Model. Cell *137*, 1005–1017.

Krek, A., Grün, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M., et al. (2005). Combinatorial microRNA target predictions. Nat. Genet. *37*, 495–500.

Krützfeldt, J., Rajewsky, N., Braich, R., Rajeev, K.G., Tuschl, T., Manoharan, M., and Stoffel, M. (2005). Silencing of microRNAs in vivo with "antagomirs". Nature *438*, 685–689.

Kundu, P., Fabian, M.R., Sonenberg, N., Bhattacharyya, S.N., and Filipowicz, W. (2012). HuR protein attenuates miRNA-mediated repression by promoting miRISC dissociation from the target RNA. Nucleic Acids Res. *40*, 5088–5100.

Laganà, A., Russo, F., Sismeiro, C., Giugno, R., Pulvirenti, A., and Ferro, A. (2010). Variability in the incidence of miRNAs and genes in fragile sites and the role of repeats and CpG islands in the distribution of genetic material. PLoS One *5*, e11166.

Lagos-Quintana, M. (2001). Identification of Novel Genes Coding for Small Expressed RNAs. Science *294*, 853–858.

Landi, M.T., Zhao, Y., Rotunno, M., Koshiol, J., Liu, H., Bergen, A.W., Rubagotti, M., Goldstein, A.M., Linnoila, I., Marincola, F.M., et al. (2010). MicroRNA expression differentiates histology and predicts survival of lung cancer. Clin. Cancer Res. *16*, 430–441.

Langenberger, D., Bermudez-Santana, C., Hertel, J., Hoffmann, S., Khaitovich, P., and Stadler, P.F. (2009). Evidence for human microRNA-offset RNAs in small RNA sequencing data. Bioinformatics *25*, 2298–2301.

Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P. a C., Monlong, J., Rivas, M. a, Gonzàlez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature *501*, 506–511.

Lau, N.C. (2001). An Abundant Class of Tiny RNAs with Probable Regulatory Roles in Caenorhabditis elegans. Science *294*, 858–862.

Lawrie, C.H., Gal, S., Dunlop, H.M., Pushkaran, B., Liggins, A.P., Pulford, K., Banham, A.H., Pezzella, F., Boultwood, J., Wainscoat, J.S., et al. (2008). Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. Br. J. Haematol. *141*, 672–675.

Lebanony, D., Benjamin, H., Gilad, S., Ezagouri, M., Dov, A., Ashkenazi, K., Gefen, N., Izraeli, S., Rechavi, G., Pass, H., et al. (2009). Diagnostic Assay Based on hsa-miR-205 Expression Distinguishes Squamous From Nonsquamous Non-Small-Cell Lung Carcinoma. J. Clin. Oncol. *27*, 2030–2037.

Lee, R.C. (2001). An Extensive Class of Small RNAs in Caenorhabditis elegans. Science *294*, 862–864.

Lee, Y.S., and Dutta, A. (2007). The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. Genes Dev. *21*, 1025–1030.

Lee, D.-C., Romero, R., Kim, J.-S., Tarca, A.L., Montenegro, D., Pineles, B.L., Kim, E., Lee, J., Kim, S.Y., Draghici, S., et al. (2011a). miR-210 Targets Iron-Sulfur Cluster Scaffold Homologue in Human Trophoblast Cell Lines. Am. J. Pathol. *179*, 590–602.

Lee, H.-C., Li, L., Gu, W., Xue, Z., Crosthwaite, S.K., Pertsemlidis, A., Lewis, Z.A., Freitag, M., Selker, E.U., Mello, C.C., et al. (2010). Diverse Pathways Generate MicroRNA-like RNAs and Dicer-Independent Small Interfering RNAs in Fungi. Mol. Cell *38*, 803–814.

Lee, J.-H., Voortman, J., Dingemans, A.-M.C., Voeller, D.M., Pham, T., Wang, Y., and Giaccone, G. (2011b). MicroRNA Expression and Clinical Outcome of Small Cell Lung Cancer. PLoS One *6*, e21300.

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

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., and Yim, J. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*.

Lee, Y., Kim, M., Han, J., Yeom, K., Lee, S., Baek, S., and Kim, V. (2004). MicroRNA genes are transcribed by RNA polymerase II. Embo J *23*, 4051–4060.

Lee, Y., Hur, I., Park, S.-Y., Kim, Y.-K., Suh, M.R., and Kim, V.N. (2006). The role of PACT in the RNA silencing pathway. EMBO J. *25*, 522–532.

Lehmann, U., Hasemeier, B., Christgen, M., Muller, M., Romermann, D., Langer, F., and Kreipe, H. (2008). Epigenetic inactivation of microRNA gene hsa-mir-9-1 in human breast cancer. J. Pathol. *214*, 17–24.

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

Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15–20.

Lin, W.C., Li, S.C., Lin, W.C., Shin, J.W., Hu, S.N., Yu, X.M., Huang, T.Y., Chen, S.C., Chen, H.C., Chen, S.J., et al. (2009). Identification of microRNA in the protist Trichomonas vaginalis. Genomics *93*, 487–493.

Liu, B., Peng, X.-C., Zheng, X.-L., Wang, J., and Qin, Y.-W. (2009). MiR-126 restoration down-regulate VEGF and inhibit the growth of lung cancer cell lines in vitro and in vivo. Lung Cancer *66*, 169–175.

Liu, C., Zhang, F., Li, T., Lu, M., Wang, L., Yue, W., and Zhang, D. (2012). MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. BMC Genomics *13*, 661.

Liu, J., Valencia-Sanchez, M.A., Hannon, G.J., and Parker, R. (2005a). MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. Nat. Cell Biol. *7*, 719–723.

Liu, J., Rivas, F. V, Wohlschlegel, J., Yates, J.R., Parker, R., and Hannon, G.J. (2005b). A role for the P-body component GW182 in microRNA function. Nat. Cell Biol. *7*, 1261–1266.

Liu, M., Shen, S., Chen, F., Yu, W., and Yu, L. (2010). Linking the septin expression with carcinogenesis. Mol. Biol. Rep. *37*, 3601–3608.

Liu, S., An, J., Lin, J., Liu, Y., Bao, L., Zhang, W., and Zhao, J.J. (2014). Single nucleotide polymorphisms of microRNA processing machinery genes and outcome of hepatocellular carcinoma. PLoS One *9*, 1–4.

Londin, E., Loher, P., Telonis, A.G., Quann, K., Clark, P., Jing, Y., Hatzimichael, E., Kirino, Y., Honda, S., Lally, M., et al. (2015). Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. Proc. Natl. Acad. Sci. *112*, 201420955.

Lu, J., and Clark, A.G. (2012). Impact of microRNA regulation on variation in human gene expression. Genome Res. *22*, 1243–1254.

Lu, J., Getz, G., Miska, E. a, Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., et al. (2005). MicroRNA expression profiles classify human cancers. Nature *435*, 834–838.

Lund, E., and Gu, S. (2004). Nuclear Export of MicroRNA. Science *303*, 95–98.

Luo, X., Yang, W., Ye, D.Q., Cui, H., Zhang, Y., Hirankarn, N., Qian, X., Tang, Y., Lau, Y.L., de Vries, N., et al. (2011). A functional variant in microRNA-146a promoter modulates its expression and confers disease risk for systemic lupus erythematosus. PLoS Genet. *7*.

Lytle, J.R., Yario, T. a, and Steitz, J. a (2007). Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. Proc. Natl. Acad. Sci. *104*, 9667–9672.

Ma, F., Xu, S., Liu, X., Zhang, Q., Xu, X., Liu, M., Hua, M., Li, N., Yao, H., and Cao, X. (2011a). The microRNA miR-29 controls innate and adaptive immune responses to intracellular bacterial infection by targeting interferon-γ. Nat. Immunol. *12*, 861–869.

Ma, L., Huang, Y., Zhu, W., Zhou, S., Zhou, J., Zeng, F., Liu, X., Zhang, Y., and Yu, J. (2011b). An Integrated Analysis of miRNA and mRNA Expressions in Non-Small Cell Lung Cancers. PLoS One *6*, e26502.

Maragkakis, M., Alexiou, P., Papadopoulos, G.L., Reczko, M., Dalamagas, T., Giannopoulos, G., Goumas, G., Koukis, E., Kourtis, K., Simossis, V.A., et al. (2009a). Accurate microRNA target prediction correlates with protein repression levels. BMC Bioinformatics *10*, 295.

Maragkakis, M., Reczko, M., Simossis, V. a., Alexiou, P., Papadopoulos, G.L., Dalamagas, T., Giannopoulos, G., Goumas, G., Koukis, E., Kourtis, K., et al. (2009b). DIANA-microT web server: Elucidating microRNA functions through target prediction. Nucleic Acids Res. *37*, 273–276.

Maragkakis, M., Alexiou, P., Papadopoulos, G.L., Reczko, M., Dalamagas, T., Giannopoulos, G., Goumas, G., Koukis, E., Kourtis, K., Simossis, V.A., et al. (2009c). Accurate microRNA target prediction correlates with protein repression levels. BMC Bioinformatics *10*, 295.

Maragkakis, M., Vergoulis, T., Alexiou, P., Reczko, M., Plomaritou, K., Gousis, M., Kourtis, K., Koziris, N., Dalamagas, T., and Hatzigeorgiou, A.G. (2011). DIANA-microT Web server upgrade supports Fly and Worm miRNA target prediction and bibliographic miRNA to disease association. Nucleic Acids Res. *39*, W145–W148.

Marco, A., MacPherson, J.I., Ronshaugen, M., and Griffiths-Jones, S. (2012). MicroRNAs from the same precursor have different targeting properties. Silence *3*, 8.

Martin, H.C., Wani, S., Steptoe, A.L., Krishnan, K., Nones, K., Nourbakhsh, E., Vlassov, A., Grimmond, S.M., and Cloonan, N. (2014). Imperfect centered miRNA

binding sites are common and can mediate repression of target mRNAs. Genome Biol. *15*, R51.

Mascaux, C., Laes, J.F., Anthoine, G., Haller, A., Ninane, V., Burny, A., and Sculier, J.P. (2009). Evolution of microRNA expression during human bronchial squamous carcinogenesis. Eur. Respir. J. *33*, 352–359.

Maute, R.L., Schneider, C., Sumazin, P., Holmes, A., Califano, A., Basso, K., and Dalla-Favera, R. (2013). tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. Proc. Natl. Acad. Sci. *110*, 1404–1409.

Mayr, C., Hemann, M.T., and Bartel, D.P. (2007). Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. Science *315*, 1576–1579.

Meiri, E., Mueller, W.C., Rosenwald, S., Zepeniuk, M., Klinke, E., Edmonston, T.B., Werner, M., Lass, U., Barshack, I., Feinmesser, M., et al. (2012). A Second-Generation MicroRNA-Based Assay for Diagnosing Tumor Tissue Origin. Oncologist *17*, 801–812.

Melo, S.A., Moutinho, C., Ropero, S., Calin, G.A., Rossi, S., Spizzo, R., Fernandez, A.F., Davalos, V., Villanueva, A., Montoya, G., et al. (2010). A genetic defect in exportin-5 traps precursor MicroRNAs in the nucleus of cancer cells. Cancer Cell *18*, 303–315.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. Nature *495*, 333–338.

Miñones-Moyano, E., Porta, S., Escaramís, G., Rabionet, R., Iraola, S., Kagerbauer, B., Espinosa-Parrilla, Y., Ferrer, I., Estivill, X., and Martí, E. (2011). MicroRNA profiling of Parkinson's disease brains identifies early downregulation of miR-34b/c which modulate mitochondrial function. Hum. Mol. Genet. *20*, 3067–3078.

Miranda, K.C., Huynh, T., Tay, Y., Ang, Y.-S., Tam, W.-L., Thomson, A.M., Lim, B., and Rigoutsos, I. (2006). A Pattern-Based Method for the Identification of MicroRNA Binding Sites and Their Corresponding Heteroduplexes. Cell *126*, 1203–1217.

Mitchell, P.S., Parkin, R.K., Kroh, E.M., Fritz, B.R., Wyman, S.K., Pogosova-Agadjanyan, E.L., Peterson, A., Noteboom, J., O'Briant, K.C., Allen, A., et al. (2008). Circulating microRNAs as stable blood-based markers for cancer detection. Proc. Natl. Acad. Sci. *105*, 10513–10518.

Mochel, F., Knight, M.A., Tong, W.-H., Hernandez, D., Ayyad, K., Taivassalo, T., Andersen, P.M., Singleton, A., Rouault, T.A., Fischbeck, K.H., et al. (2008). Splice Mutation in the Iron-Sulfur Cluster Scaffold Protein ISCU Causes Myopathy with Exercise Intolerance. Am. J. Hum. Genet. *82*, 652–660.

Molina, J.R., Yang, P., Cassivi, S.D., Schild, S.E., and Adjei, A. a. (2008). Non-Small Cell Lung Cancer: Epidemiology, Risk Factors, Treatment, and Survivorship. Mayo Clin. Proc. *83*, 584–594.

Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.-L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., et al. (2008). Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res. *18*, 610–621.

Nadal, E., Truini, A., Nakata, A., Lin, J., Reddy, R.M., Chang, A.C., Ramnath, N., Gotoh, N., Beer, D.G., and Chen, G. (2015). A Novel Serum 4-microRNA Signature for Lung Cancer Detection. Sci. Rep. *5*, 12464.

Nicoloso, M.S., Sun, H., Spizzo, R., Kim, H., Wickramasinghe, P., Shimizu, M., Wojcik, S.E., Ferdin, J., Kunej, T., Xiao, L., et al. (2010). Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. Cancer Res. *70*, 2789–2798.

Nymark, P., Guled, M., Borze, I., Faisal, A., Lahti, L., Salmenkivi, K., Kettunen, E., Anttila, S., and Knuutila, S. (2011). Integrative analysis of microRNA, mRNA and aCGH data reveals asbestos- and histology-related changes in lung cancer. Genes, Chromosom. Cancer *50*, 585–597.

O'Donnell, K. a, Wentzel, E. a, Zeller, K.I., Dang, C. V, and Mendell, J.T. (2005). c-Myc-regulated microRNAs modulate E2F1 expression. Nature *435*, 839–843.

Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C. (2007). The Mirtron Pathway Generates microRNA-Class Regulatory RNAs in Drosophila. Cell *130*, 89–100.

Okamura, K., Phillips, M.D., Tyler, D.M., Duan, H., Chou, Y., and Lai, E.C. (2008). The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution. Nat. Struct. Mol. Biol. *15*, 354–363.

Olson, P., Lu, J., Zhang, H., Shai, A., Chun, M.G., Wang, Y., Libutti, S.K., Nakakura, E.K., Golub, T.R., and Hanahan, D. (2009). MicroRNA dynamics in the stages of tumorigenesis correlate with hallmark capabilities of cancer. Genes Dev. *23*, 2152–2165.

Orom, U.A., Nielsen, F.C., and Lund, A.H. (2008). MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. Mol Cell *30*.

Paczynska, P., Grzemski, A., and Szydlowski, M. (2015). Distribution of miRNA genes in the pig genome. BMC Genet. *16*, 6.

Paraskevopoulou, M.D., Georgakilas, G., Kostoulas, N., Vlachos, I.S., Vergoulis, T., Reczko, M., Filippidis, C., Dalamagas, T., and Hatzigeorgiou, A.G. (2013). DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. Nucleic Acids Res. *41*, W169–W173.

Pasquinelli, a E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Müller, P., et al. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature *408*, 86–89.

Patnaik, S., Mallick, R., Kannisto, E., Sharma, R., Bshara, W., Yendamuri, S., and Dhillon, S.S. (2015). MiR-205 and MiR-375 microRNA assays to distinguish squamous cell carcinoma from adenocarcinoma in lung cancer biopsies. J. Thorac. Oncol. *10*, 446–453.

Petersen, C.P., Bordeleau, M.E., Pelletier, J., and Sharp, P.A. (2006). Short RNAs repress translation after initiation in mammalian cells. Mol. Cell *21*, 533–542.

Pigati, L., Yaddanapudi, S.C.S., Iyengar, R., Kim, D.J., Hearn, S.A., Danforth, D., Hastings, M.L., and Duelli, D.M. (2010). Selective release of MicroRNA species from normal and malignant mammary epithelial cells. PLoS One *5*.

Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E., and Filipowicz, W. (2005). Inhibition of Translational Initiation by Let-7 MicroRNA in Human Cells. Science *309*, 1573–1576.

Place, R.F., Li, L.-C., Pookot, D., Noonan, E.J., and Dahiya, R. (2008). MicroRNA-373 induces expression of genes with complementary promoter sequences. Proc. Natl. Acad. Sci. *105*, 1608–1613.

Puisségur, M.-P., Mazure, N.M., Bertero, T., Pradelli, L., Grosso, S., Robbe-Sermesant, K., Maurin, T., Lebrigand, K., Cardinaud, B., Hofman, V., et al. (2011). miR-210 is

overexpressed in late stages of lung cancer and mediates mitochondrial alterations associated with modulation of HIF-1 activity. Cell Death Differ. *18*, 465–478.
Purrington, K.S., Slager, S., Eccles, D., Yannoukakos, D., Fasching, P.A., Miron, P., Carpenter, J., Chang-Claude, J., Martin, N.G., Montgomery, G.W., et al. (2014). Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer. Carcinogenesis *35*, 1012–1019.
Rajewsky, N. (2006). microRNA target predictions in animals. Nat. Genet. *38*, S8–S13.
Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C.E., Socci, N.D., and Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome Biol. *14*, R95.
Raponi, M., Dossey, L., Jatkoe, T., Wu, X., Chen, G., Fan, H., and Beer, D.G. (2009). MicroRNA classifiers for predicting prognosis of squamous cell lung cancer. Cancer Res. *69*, 5776–5783.
Rawlings-Goss, R.A., Campbell, M.C., and Tishkoff, S.A. (2014). Global population-specific variation in miRNA associated with cancer risk and clinical biomarkers. BMC Med. Genomics *7*, 53.
Reczko, M., Maragkakis, M., Alexiou, P., Grosse, I., and Hatzigeorgiou, A.G. (2012). Functional microRNA targets in protein coding sequences. Bioinformatics *28*, 771–776.
Rehwinkel, J., Behm-Ansmant, I., and Gatfield, D. (2005). miRNA-mediated gene silencing A crucial role for GW182 and the DCP1 : DCP2 decapping complex in miRNA-mediated gene silencing. RNA *11*, 1640–1647.
Reinhart, B.J. (2002). MicroRNAs in plants. Genes Dev. *16*, 1616–1626.
Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901–906.
Richardson, K., Lai, C.-Q., Parnell, L.D., Lee, Y.-C., and Ordovas, J.M. (2011). A genome-wide survey for SNPs altering microRNA seed sites identifies functional candidates in GWAS. BMC Genomics *12*, 504.
Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.
Rodriguez, A. (2004). Identification of Mammalian microRNA Host Genes and Transcription Units. Genome Res. *14*, 1902–1910.
van Rooij, E., Sutherland, L.B., Liu, N., Williams, A.H., McAnally, J., Gerard, R.D., Richardson, J. a, and Olson, E.N. (2006). A signature pattern of stress-responsive microRNAs that can evoke cardiac hypertrophy and heart failure. Proc. Natl. Acad. Sci. *103*, 18255–18260.
Rosenfeld, N., Aharonov, R., Meiri, E., Rosenwald, S., Spector, Y., Zepeniuk, M., Benjamin, H., Shabes, N., Tabak, S., Levy, A., et al. (2008). MicroRNAs accurately identify cancer tissue origin. Nat. Biotechnol. *26*, 462–469.
Roth, C., Rack, B., Müller, V., Janni, W., Pantel, K., and Schwarzenbach, H. (2010). Circulating microRNAs as blood-based markers for patients with primary and metastatic breast cancer. Breast Cancer Res. *12*, R90.
Ruby, J.G., Jan, C.H., and Bartel, D.P. (2007). Intronic microRNA precursors that bypass Drosha processing. Nature *448*, 83–86.
Salzman, J., Gawad, C., Wang, P.L., Lacayo, N., and Brown, P.O. (2012). Circular RNAs Are the Predominant Transcript Isoform from Hundreds of Human Genes in Diverse Cell Types. PLoS One *7*, e30733.

Sampson, V.B., Rong, N.H., Han, J., Yang, Q., Aris, V., Soteropoulos, P., Petrelli, N.J., Dunn, S.P., and Krueger, L.J. (2007). MicroRNA let-7a down-regulates MYC and reverts MYC-induced growth in Burkitt lymphoma cells. Cancer Res. *67*, 9762–9770.

Saunders, M.A., Liang, H., and Li, W.-H. (2007). Human polymorphism at microRNAs and microRNA target sites. Proc. Natl. Acad. Sci. *104*, 3300–3305.

Schwarz, D.S., Hutvágner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199–208.

Seike, M., Goto, A., Okano, T., Bowman, E.D., Schetter, A.J., Horikawa, I., Mathe, E.A., Jen, J., Yang, P., Sugimura, H., et al. (2009). MiR-21 is an EGFR-regulated anti-apoptotic factor in lung cancer in never-smokers. Proc Natl Acad Sci. *106*, 12085–12090.

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. Nature *455*, 58–63.

Sethupathy, P., Megraw, M., and Hatzigeorgiou, A.G. (2006). A guide through present computational approaches for the identification of mammalian microRNA targets. Nat. Methods *3*, 881–886.

Sevignani, C., Calin, G. a, Nnadi, S.C., Shimizu, M., Davuluri, R. V, Hyslop, T., Demant, P., Croce, C.M., and Siracusa, L.D. (2007). MicroRNA genes are frequently located near mouse cancer susceptibility loci. Proc. Natl. Acad. Sci. *104*, 8017–8022.

Shell, S., Park, S.-M., Radjabi, A.R., Schickel, R., Kistner, E.O., Jewell, D. a, Feig, C., Lengyel, E., and Peter, M.E. (2007). Let-7 expression defines two differentiation stages of cancer. Proc. Natl. Acad. Sci. *104*, 11400–11405.

Shi, W., Hendrix, D., Levine, M., and Haley, B. (2009). A distinct class of small RNAs arises from pre-miRNA-proximal regions in a simple chordate. Nat Struct Mol Biol *16*, 183–189.

Simone, N.L., Soule, B.P., Ly, D., Saleh, A.D., Savage, J.E., DeGraff, W., Cook, J., Harris, C.C., Gius, D., and Mitchell, J.B. (2009). Ionizing radiation-induced oxidative stress alters miRNA expression. PLoS One *4*, 1–7.

Smyth, G.K. (2004). Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. Stat. Appl. Genet. Mol. Biol. *3*, 1–26.

Son, J.W., Kim, Y.J., Cho, H.M., and Lee, S.Y. (2009). MicroRNA Expression Profiles in Korean Non-Small Cell Lung Cancer. Tuberc. Respir. Dis. (Seoul). *67*, 413–421.

Soneson, C., and Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics *14*, 91.

Stark, A., Brennecke, J., Russell, R.B., and Cohen, S.M. (2003). Identification of Drosophila MicroRNA Targets. PLoS Biol. *1*, e60.

Stark, A., Brennecke, J., Bushati, N., Russell, R.B., and Cohen, S.M. (2005). Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. Cell *123*, 1133–1146.

Sun, C., Liu, Z., Li, S., Yang, C., Xue, R., Xi, Y., Wang, L., Wang, S., He, Q., Huang, J., et al. (2015). Down-regulation of c-Met and Bcl2 by microRNA-206, activates apoptosis, and inhibits tumor cell proliferation, migration and colony formation. Oncotarget *6*, 25533–25574.

Sun, L., Goff, L. a, Trapnell, C., Alexander, R., Lo, K.A., Hacisuleyman, E., Sauvageau, M., Tazon-Vega, B., Kelley, D.R., Hendrickson, D.G., et al. (2013). Long noncoding RNAs regulate adipogenesis. Proc. Natl. Acad. Sci. *110*, 3387–3392.

Sun, Y., Bai, Y., Zhang, F., Wang, Y., Guo, Y., and Guo, L. (2010). miR-126 inhibits non-small cell lung cancer cells proliferation by targeting EGFL7. Biochem. Biophys. Res. Commun. *391*, 1483–1489.

Takamizawa, J. (2004). Reduced Expression of the let-7 MicroRNAs in Human Lung Cancers in Association with Shortened Postoperative Survival. Cancer Res. *64*, 3753–3756.

Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., et al. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature *453*, 534–538.

Tay, Y., Zhang, J., Thomson, A.M., Lim, B., and Rigoutsos, I. (2008). MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. Nature *455*, 1124–1128.

Tazawa, H., Tsuchiya, N., Izumiya, M., and Nakagama, H. (2007). Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells. Proc Natl Acad Sci. *104*.

Thomas, L.F., Saito, T., and Saetrom, P. (2011). Inferring causative variants in microRNA target sites. Nucleic Acids Res. *39*, e109–e109.

Thorgeirsson, T.E., Geller, F., Sulem, P., Rafnar, T., Wiste, A., Magnusson, K.P., Manolescu, A., Thorleifsson, G., Stefansson, H., Ingason, A., et al. (2008). A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature *452*, 638–642.

Vaishnavi, V., Manikandan, M., and Munirajan, A.K. (2014). Mining the 3′UTR of Autism-implicated Genes for SNPs Perturbing MicroRNA Regulation. Genomics. Proteomics Bioinformatics *12*, 92–104.

Valadi, H., Ekström, K., Bossios, A., Sjöstrand, M., Lee, J.J., and Lötvall, J.O. (2007). Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nat. Cell Biol. *9*, 654–659.

Välk, K., Vooder, T., Kolde, R., Reintam, M.-A., Petzold, C., Vilo, J., and Metspalu, A. (2010). Gene expression profiles of non-small cell lung cancer: survival prediction and new biomarkers. Oncology *79*, 283–292.

Vasudevan, S., Tong, Y., and Steitz, J. a (2007). Switching from repression to activation: microRNAs can up-regulate translation. Science *318*, 1931–1934.

Del Vescovo, V., Grasso, M., Barbareschi, M., and Denti, M.A. (2014). MicroRNAs as lung cancer biomarkers. World J. Clin. Oncol. *5*, 604–620.

Vickers, K.C., Palmisano, B.T., Shoucri, B.M., Shamburek, R.D., and Remaley, A.T. (2011). MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. Nat. Cell Biol. *13*, 423–433.

Vlachos, I.S., Paraskevopoulou, M.D., Karagkouni, D., Georgakilas, G., Vergoulis, T., Kanellos, I., Anastasopoulos, I.-L., Maniou, S., Karathanou, K., Kalfakakou, D., et al. (2015). DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. Nucleic Acids Res. *43*, D153–D159.

Volinia, S., Calin, G. a, Liu, C.-G., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M., et al. (2006). A microRNA expression signature of human solid tumors defines cancer gene targets. Proc. Natl. Acad. Sci. *103*, 2257–2261.

Võsa, U., Vooder, T., Kolde, R., Fischer, K., Välk, K., Tõnisson, N., Roosipuu, R., Vilo, J., Metspalu, A., and Annilo, T. (2011). Identification of miR-374a as a prognostic marker for survival in patients with early-stage nonsmall cell lung cancer. Genes, Chromosom. Cancer *50*, 812–822.

Wahlquist, C., Jeong, D., Rojas-Muñoz, A., Kho, C., Lee, A., Mitsuyama, S., van Mil, A., Park, W.J., Sluijter, J.P.G., Doevendans, P. a F., et al. (2014). Inhibition of miR-25 improves cardiac contractility in the failing heart. Nature *508*, 531–535.

Wang, X. (2014). Composition of seed sequence is a major determinant of microRNA targeting patterns. Bioinformatics *30*, 1377–1383.

Wang, X., and El Naqa, I.M. (2008). Prediction of both conserved and nonconserved microRNA targets in animals. Bioinformatics *24*, 325–332.

Wang, C., Ding, M., Xia, M., Chen, S., Van Le, A., Soto-Gil, R., Shen, Y., Wang, N., Wang, J., Gu, W., et al. (2015). A Five-miRNA Panel Identified From a Multicentric Case–control Study Serves as a Novel Diagnostic Tool for Ethnically Diverse Non-small-cell Lung Cancer Patients. EBioMedicine *2*, 1–9.

Wang, J., Bansal, a T., Martin, M., Germer, S., Benayed, R., Essioux, L., Lee, J.S., Begovich, a, Hemmings, a, Kenwright, a, et al. (2013). Genome-wide association analysis implicates the involvement of eight loci with response to tocilizumab for the treatment of rheumatoid arthritis. Pharmacogenomics J. *13*, 235–241.

Wang, R., Wang, Z.-X., Yang, J.-S., Pan, X., De, W., and Chen, L.-B. (2011). MicroRNA-451 functions as a tumor suppressor in human non-small cell lung cancer by targeting ras-related protein 14 (RAB14). Oncogene *30*, 2644–2658.

Wang, Y., Xia, H., Zhuang, Z., Miao, L., Chen, X., and Cai, H. (2014). Axl-altered microRNAs regulate tumorigenicity and gefitinib resistance in lung cancer. Cell Death Dis. *5*, e1227.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., et al. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature *453*, 539–543.

Weber, J.A., Baxter, D.H., Zhang, S., Huang, D.Y., Huang, K.H., Lee, M.J., Galas, D.J., and Wang, K. (2010). The microRNA spectrum in 12 body fluids. Clin. Chem. *56*, 1733–1741.

Wei, R., Yang, F., Urban, T.J., Li, L., Chalasani, N., Flockhart, D.A., and Liu, W. (2012). Impact of the Interaction between 3′-UTR SNPs and microRNA on the Expression of Human Xenobiotic Metabolism Enzyme and Transporter Genes. Front. Genet. *3*, 1–11.

Westra, H.-J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. *45*, 1238–1243.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell *75*, 855–862.

Wu, C., Kraft, P., Zhai, K., Chang, J., Wang, Z., Li, Y., Hu, Z., He, Z., Jia, W., Abnet, C.C., et al. (2012). Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. Nat. Genet. *44*, 1090–1097.

Wu, M., Jolicoeur, N., Li, Z., Zhang, L., Fortin, Y., L'Abbe, D., Yu, Z., and Shen, S.H. (2008). Genetic variations of microRNAs in human cancer and their effects on the expression of miRNAs. Carcinogenesis *29*, 1710–1716.

Wu, S., Jia, S., and Xu, P. (2014). MicroRNA-9 as a novel prognostic biomarker in human laryngeal squamous cell carcinoma. Int. J. Clin. Exp. Med. *7*, 5523–5528.

Wynendaele, J., Bohnke, A., Leucci, E., Nielsen, S.J., Lambertz, I., Hammer, S., Sbrzesny, N., Kubitza, D., Wolf, A., Gradhand, E., et al. (2010). An Illegitimate microRNA Target Site within the 3' UTR of MDM4 Affects Ovarian Cancer Progression and Chemosensitivity. Cancer Res. *70*, 9641–9649.

Xiao, C., Srinivasan, L., Calado, D.P., Patterson, H.C., Zhang, B., Wang, J., Henderson, J.M., Kutok, J.L., and Rajewsky, K. (2008). Lymphoproliferative disease and autoimmunity in mice with increased miR-17-92 expression in lymphocytes. Nat. Immunol. *9*, 405–414.

Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., and Li, T. (2009). miRecords: an integrated resource for microRNA-target interactions. Nucleic Acids Res. *37*, D105–D110.

Xie, Y., Todd, N.W., Liu, Z., Zhan, M., Fang, H., Peng, H., Alattar, M., Deepak, J., Stass, S. a, and Jiang, F. (2010). Altered miRNA expression in sputum for diagnosis of non-small cell lung cancer. Lung Cancer *67*, 170–176.

Xing, L., Todd, N.W., Yu, L., Fang, H., and Jiang, F. (2010). Early detection of squamous cell lung cancer in sputum by a panel of microRNA markers. Mod. Pathol. *23*, 1157–1164.

Xu, T., Liu, X., Han, L., Shen, H., Liu, L., and Shu, Y. (2014). Up-regulation of miR-9 expression as a poor prognostic biomarker in patients with non-small cell lung cancer. Clin. Transl. Oncol. *16*, 469–475.

Xu, X., Wang, W., Su, N., Zhu, X., Yao, J., Gao, W., Hu, Z., and Sun, Y. (2015). miR-374a promotes cell proliferation, migration and invasion by targeting SRCIN1 in gastric cancer. FEBS Lett. *589*, 407–413.

Yanaihara, N., Caplen, N., Bowman, E., Seike, M., Kumamoto, K., Yi, M., Stephens, R.M., Okamoto, A., Yokota, J., Tanaka, T., et al. (2006). Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell *9*, 189–198.

Yang, J.S., and Lai, E.C. (2010). Dicer-independent, Ago2-mediated microRNA biogenesis in vertebrates. Cell Cycle *9*, 4455–4460.

Yang, Y., Li, X., Yang, Q., Wang, X., Zhou, Y., Jiang, T., Ma, Q., and Wang, Y.-J. (2010). The role of microRNA in human lung squamous cell carcinoma. Cancer Genet. Cytogenet. *200*, 127–133.

Ye, Y., Wang, K.K., Gu, J., Yang, H., Lin, J., Ajani, J.A., and Wu, X. (2008). Genetic variations in MicroRNA-related genes are novel susceptibility loci for esophageal cancer risk. Cancer Prev. Res. *1*, 460–469.

Yekta, S. (2004). MicroRNA-Directed Cleavage of HOXB8 mRNA. Science *304*, 594–596.

Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev. *17*, 3011–3016.

Yoda, M., Cifuentes, D., Izumi, N., Sakaguchi, Y., Suzuki, T., Giraldez, A.J., and Tomari, Y. (2013). Poly(A)-specific ribonuclease mediates 3'-end trimming of argonaute2-cleaved precursor microrna. Cell Rep. *5*, 715–726.

Yu, L., Todd, N.W., Xing, L., Xie, Y., Zhang, H., Liu, Z., Fang, H., Zhang, J., Katz, R.L., and Jiang, F. (2010). Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers. Int. J. Cancer *127*, 2870–2878.

Zernecke, A., Bidzhekov, K., Noels, H., Shagdarsuren, E., Gan, L., Denecke, B., Hristov, M., Köppel, T., Jahantigh, M.N., Lutgens, E., et al. (2009). Delivery of microRNA-126 by apoptotic bodies induces CXCL12-dependent vascular protection. Sci. Signal. *2*, ra81.

Zhang, J. guang, Wang, J. jun, Zhao, F., Liu, Q., Jiang, K., and Yang, G. hai (2010). MicroRNA-21 (miR-21) represses tumor suppressor PTEN and promotes growth and invasion in non-small cell lung cancer (NSCLC). Clin. Chim. Acta *411*, 846–852.

Zhang, W., Edwards, A., Zhu, D., Flemington, E.K., Deininger, P., and Zhang, K. (2012a). miRNA-Mediated Relationships between Cis-SNP Genotypes and Transcript Intensities in Lymphocyte Cell Lines. PLoS One *7*, e31429.

Zhang, Y., Geng, L., Talmon, G., and Wang, J. (2015a). MicroRNA-520g Confers Drug Resistance by Regulating p21 Expression in Colorectal Cancer. J. Biol. Chem. *290*, 6215–6225.

Zhang, Y.J., Xu, F., Zhang, Y.J., Li, H.B., Han, J.C., and Li, L. (2015b). miR-206 inhibits non small cell lung cancer cell proliferation and invasion by targeting SOX9. Int J Clin Exp Med *8*, 9107–9113.

Zhang, Y.K., Zhu, W.Y., He, J.Y., Chen, D.D., Huang, Y.Y., Le, H.B., and Liu, X.G. (2012b). MiRNAs expression profiling to distinguish lung squamous-cell carcinoma from adenocarcinoma subtypes. J. Cancer Res. Clin. Oncol. *138*, 1641–1650.

Zhernakova, D. V, de Klerk, E., Westra, H.-J., Mastrokolias, A., Amini, S., Ariyurek, Y., Jansen, R., Penninx, B.W., Hottenga, J.J., Willemsen, G., et al. (2013). DeepSAGE reveals genetic variants associated with alternative polyadenylation and expression of coding and non-coding transcripts. PLoS Genet. *9*, e1003594.

Zhou, H., Arcila, M.L., Li, Z., Lee, E.J., Henzler, C., Liu, J., Rana, T.M., and Kosik, K.S. (2012). Deep annotation of mouse iso-miR and iso-moR variation. Nucleic Acids Res. *40*, 5864–5875.

Zu, Y., Ban, J., Xia, Z., Wang, J., Cai, Y., Ping, W., and Sun, W. (2013). Genetic variation in a miR-335 binding site in BIRC5 alters susceptibility to lung cancer in Chinese Han populations. Biochem. Biophys. Res. Commun. *430*, 529–534.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my supervisor Tarmo Annilo for introducing me the “noncoding world”, as well as for his guidance and support over all these years. I would like to thank Prof. Metspalu for providing the good scientific working atmosphere, and for the inspiration and motivation.

I am grateful for all of my colleagues at the Estonian Genome Center, the senior scientists and scientists for their valuable feedback and expertise, my fellow PhD students for their friendly support on this bumpy road we are all on, and all of the administrative staff (especially Krista Liiv) for their help with practical matters. I thank Viljo Soo for excellent work on generating the expression data used in this study, and all of the other lab technicians for their help while I was still working in the wet lab.

I thank Tõnu Vooder for making it possible to work with cancer samples, and for his essential contribution to the research presented in this thesis. I thank Raivo Kolde for providing me with the first real analysis scripts and sparking my interest in the analysis of biological data. I would also like to thank Evelin Mihhailov and Tõnu Esko for their expertise and help with genotype data. I thank all of the coauthors of the papers presented in this study for their contributions and help.

During my studies, I had an opportunity to intern in the Genetics Department of the University of Groningen. I would like to thank Prof. Cisca Wijmenga for granting me this wonderful opportunity, Vinod Kumar and Sebo Withoff for their great supervision, and my fellow PhD students for making me feel welcome.

I would like to thank my current colleagues, Marc Jan Bonder, Patrick Deelen, Dasha Zhernakova, and Adriaan van der Graaf, for their ideas and suggestions for the last part of the thesis. In particular, I would like to thank Prof. Lude Franke for his input and the time he provided me to finish this thesis.

Last but not least, I would like to thank my family and friends for their moral support on this long journey. My special gratitude goes to Mariann – it would have been much harder without your unconditional support, belief, and encouragement.

This research was financially supported by EU FP7 grants 306031, 313010, EU H2020 grants 692145, 676550, 654248, the European Regional Development Fund, Center of Excellence in Genomics (EXCEGEN) and University of Tartu (SP1GVARENG), Estonian Research Council Grant IUT20-60, IUT24-6, TerVE programme grant PerMed I. I acknowledge The Graduate School in Biomedicine and Biotechnology and DoRa Programme for financing my attendance to several conferences.

PUBLICATIONS

CURRICULUM VITAE

Name: Urmo Võsa
Date of birth: January 29, 1986
Citizenship: Estonian
Contact: University of Tartu, The Institute of Molecular and Cell Biology, Department of Biotechnology, Riia 23, 51010, Tartu
E-mail: urmo.vosa@ut.ee

Education:

2010–...	University of Tartu, PhD student (Molecular- ja Cell Biology)
2008–2010	University of Tartu, MSc (Biology)
2007–2008	Service in Estonian Defence Forces
2004–2007	University of Tartu, BSc (Biology)
2001–2004	Viljandi C. R. Jakobson Gymnasium

Working experience:

2015–...	Postdoctoral researcher, University of Groningen
2012–2015	Specialist, Estonian Genome Center, University of Tartu

Publications:

Võsa, U., Esko, T., Kasela, S., and Annilo, T. (2015). Altered gene expression associated with microRNA binding site polymorphisms. PLoS One 10, e0141351.

Krjutškov, K., Koltšina, M., Grand, K., **Võsa, U.**, Sauk, M., Tõnisson, N., and Salumets, A. (2014). Tissue-specific mitochondrial heteroplasmy at position 16,093 within the same individual. Curr. Genet. 60, 11–16.

Kumar, V., Westra, H.J., Karjalainen, J., Zhernakova, D. V., Esko, T., Hrdlickova, B., Almeida, R., Zhernakova, A., Reinmaa, E., **Võsa, U.**, et al. (2013). Human Disease-Associated Genetic Variation Impacts Large Intergenic Non-Coding RNA Expression. PLoS Genet. 9, e1003201.

Tammiste, A., Jiang, T., Fischer, K., Magi, R., Krjutskov, K., Pettai, K., Esko, T., Li, Y., Tansey, K.E., Carroll, L.S., Uher, R, McGuffin, P, **Võsa, U.**, et al. (2013). Whole-exome sequencing identifies a polymorphism in the BMP5 gene associated with SSRI treatment response in major depression. J Psychopharmacol 27, 915–920.

Võsa, U., Vooder, T., Kolde, R., Vilo, J., Metspalu, A., and Annilo, T. (2013). Meta-analysis of microRNA expression in lung cancer. Int. J. Cancer 132, 2884–2893.

Lokk, K., Vooder, T., Kolde, R., Välk, K., **Võsa, U.**, Roosipuu, R., Milani, L., Fischer, K., Koltsina, M., Urgard, E., et al. (2012). Methylation markers of early-stage non-small cell lung cancer. PLoS One 7, e39813.

Võsa, U., Vooder, T., Kolde, R., Fischer, K., Välk, K., Tõnisson, N., Roosipuu, R., Vilo, J., Metspalu, A., and Annilo, T. (2011). Identification of miR-374a as a prognostic marker for survival in patients with early-stage nonsmall cell lung cancer. Genes, Chromosom. Cancer 50, 812–822.

Urgard, E., Vooder, T., **Võsa, U.**, Välk, K., Liu, M., Luo, C., Hoti, F., Roosipuu, R., Annilo, T., Laine, J., et al. (2011). Metagenes Associated with Survival in Non-Small Cell Lung Cancer. Cancer Inform. 10, 1–9.

Supervised dissertations:

Co-supervision of the bachelor thesis of Klaarika Kikas "Effects of polymorphisms in mir-*196a-2*, *KRAS* and *TERT* genes on cancer susceptibility and survival" ("Ühenukleotiidsete polümorfismide seos kopsuvähi riski ja patsientide elulemusega *mir-196a-2*, *KRAS* ja *TERT* geeni näitel") (2013)

Organisatorial membership:

Estonian Seminatural Community Conservation Association
Estonian Naturalists Society
Estonian Defence League

ELULOOKIRJELDUS

Nimi: Urmo Võsa
Sünniaeg: 29. jaanuar, 1986
Kodakondsus: Eesti
Aadress: Tartu Ülikool, Molekulaar- ja Rakubioloogia Instituut, Biotehnoloogia õppetool, Riia 23, 51010, Tartu
E-post: urmo.vosa@ut.ee

Haridus:

2010–...	Tartu Ülikool, doktoriõpe (molekulaar- ja rakubioloogia)
2008–2010	Tartu Ülikool, loodusteaduste magistri kraad (bioloogia)
2007–2008	Ajateenistus Eesti Kaitseväes
2004–2007	Tartu Ülikool, loodusteaduste bakalaureuse kraad (bioloogia)
2001–2004	Viljandi C. R. Jakobsoni nim. Gümnaasium (keskharidus)

Töökogemus:

2015–...	Teadur, Groningeni Ülikool
2012–2015	Spetsialist, Tartu Ülikooli Eesti Geenivaramu

Teaduspublikatsioonid:

Võsa, U., Esko, T., Kasela, S., and Annilo, T. (2015). Altered gene expression associated with microRNA binding site polymorphisms. PLoS One 10, e0141351.

Krjutškov, K., Koltšina, M., Grand, K., **Võsa, U.**, Sauk, M., Tõnisson, N., and Salumets, A. (2014). Tissue-specific mitochondrial heteroplasmy at position 16,093 within the same individual. Curr. Genet. 60, 11–16.

Kumar, V., Westra, H.J., Karjalainen, J., Zhernakova, D. V., Esko, T., Hrdlickova, B., Almeida, R., Zhernakova, A., Reinmaa, E., **Võsa, U.**, et al. (2013). Human Disease-Associated Genetic Variation Impacts Large Intergenic Non-Coding RNA Expression. PLoS Genet. 9, e1003201.

Tammiste, A., Jiang, T., Fischer, K., Magi, R., Krjutskov, K., Pettai, K., Esko, T., Li, Y., Tansey, K.E., Carroll, L.S., Uher, R, McGuffin, P, **Võsa, U.**, et al. (2013). Whole-exome sequencing identifies a polymorphism in the BMP5 gene associated with SSRI treatment response in major depression. J Psychopharmacol 27, 915–920.

Võsa, U., Vooder, T., Kolde, R., Vilo, J., Metspalu, A., and Annilo, T. (2013). Meta-analysis of microRNA expression in lung cancer. Int. J. Cancer 132, 2884–2893.

Lokk, K., Vooder, T., Kolde, R., Välk, K., **Võsa, U.**, Roosipuu, R., Milani, L., Fischer, K., Koltsina, M., Urgard, E., et al. (2012). Methylation markers of early-stage non-small cell lung cancer. PLoS One 7, e39813.

Võsa, U., Vooder, T., Kolde, R., Fischer, K., Välk, K., Tõnisson, N., Roosipuu, R., Vilo, J., Metspalu, A., and Annilo, T. (2011). Identification of miR-374a

as a prognostic marker for survival in patients with early-stage nonsmall cell lung cancer. Genes, Chromosom. Cancer 50, 812–822.
Urgard, E., Vooder, T., **Võsa, U.**, Välk, K., Liu, M., Luo, C., Hoti, F., Roosipuu, R., Annilo, T., Laine, J., et al. (2011). Metagenes Associated with Survival in Non-Small Cell Lung Cancer. Cancer Inform. 10, 1–9.

Juhendatud väitekirjad:
Kaasjuhendamine Klaarika Kikas'e bakalaureusetööle "Ühenukleotiidsete polümorfismide seos kopsuvähi riski ja patsientide elulemusega *mir-196a-2*, *KRAS* ja *TERT* geeni näitel" ("Effects of polymorphisms in *mir-196a-2*, *KRAS* and *TERT* genes on cancer susceptibility and survival") (2013)

Kuulumine organisatsioonidesse:
Pärandkoosluste Kaitse Ühing
Eesti Looduseuurijate Selts
Kaitseliit

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS

1. **Toivo Maimets**. Studies of human oncoprotein p53. Tartu, 1991, 96 p.
2. **Enn K. Seppet**. Thyroid state control over energy metabolism, ion transport and contractile functions in rat heart. Tartu, 1991, 135 p.
3. **Kristjan Zobel**. Epifüütsete makrosamblike väärtus õhu saastuse indikaatoritena Hamar-Dobani boreaalsetes mägimetsades. Tartu, 1992, 131 lk.
4. **Andres Mäe**. Conjugal mobilization of catabolic plasmids by transposable elements in helper plasmids. Tartu, 1992, 91 p.
5. **Maia Kivisaar**. Studies on phenol degradation genes of *Pseudomonas* sp. strain EST 1001. Tartu, 1992, 61 p.
6. **Allan Nurk**. Nucleotide sequences of phenol degradative genes from *Pseudomonas sp.* strain EST 1001 and their transcriptional activation in *Pseudomonas putida.* Tartu, 1992, 72 p.
7. **Ülo Tamm**. The genus *Populus* L. in Estonia: variation of the species biology and introduction. Tartu, 1993, 91 p.
8. **Jaanus Remme**. Studies on the peptidyltransferase centre of the *E.coli* ribosome. Tartu, 1993, 68 p.
9. **Ülo Langel**. Galanin and galanin antagonists. Tartu, 1993, 97 p.
10. **Arvo Käärd**. The development of an automatic online dynamic fluorescense-based pH-dependent fiber optic penicillin flowthrought biosensor for the control of the benzylpenicillin hydrolysis. Tartu, 1993, 117 p.
11. **Lilian Järvekülg**. Antigenic analysis and development of sensitive immunoassay for potato viruses. Tartu, 1993, 147 p.
12. **Jaak Palumets**. Analysis of phytomass partition in Norway spruce. Tartu, 1993, 47 p.
13. **Arne Sellin**. Variation in hydraulic architecture of *Picea abies* (L.) Karst. trees grown under different enviromental conditions. Tartu, 1994, 119 p.

13. **Mati Reeben**. Regulation of light neurofilament gene expression. Tartu, 1994, 108 p.
14. **Urmas Tartes**. Respiration rhytms in insects. Tartu, 1995, 109 p.
15. **Ülo Puurand**. The complete nucleotide sequence and infections *in vitro* transcripts from cloned cDNA of a potato A potyvirus. Tartu, 1995, 96 p.
16. **Peeter Hõrak**. Pathways of selection in avian reproduction: a functional framework and its application in the population study of the great tit (*Parus major*). Tartu, 1995, 118 p.
17. **Erkki Truve**. Studies on specific and broad spectrum virus resistance in transgenic plants. Tartu, 1996, 158 p.
18. **Illar Pata**. Cloning and characterization of human and mouse ribosomal protein S6-encoding genes. Tartu, 1996, 60 p.
19. **Ülo Niinemets**. Importance of structural features of leaves and canopy in determining species shade-tolerance in temperature deciduous woody taxa. Tartu, 1996, 150 p.

20. **Ants Kurg**. Bovine leukemia virus: molecular studies on the packaging region and DNA diagnostics in cattle. Tartu, 1996, 104 p.
21. **Ene Ustav**. E2 as the modulator of the BPV1 DNA replication. Tartu, 1996, 100 p.
22. **Aksel Soosaar**. Role of helix-loop-helix and nuclear hormone receptor transcription factors in neurogenesis. Tartu, 1996, 109 p.
23. **Maido Remm**. Human papillomavirus type 18: replication, transformation and gene expression. Tartu, 1997, 117 p.
24. **Tiiu Kull**. Population dynamics in *Cypripedium calceolus* L. Tartu, 1997, 124 p.
25. **Kalle Olli**. Evolutionary life-strategies of autotrophic planktonic microorganisms in the Baltic Sea. Tartu, 1997, 180 p.
26. **Meelis Pärtel**. Species diversity and community dynamics in calcareous grassland communities in Western Estonia. Tartu, 1997, 124 p.
27. **Malle Leht**. The Genus *Potentilla* L. in Estonia, Latvia and Lithuania: distribution, morphology and taxonomy. Tartu, 1997, 186 p.
28. **Tanel Tenson**. Ribosomes, peptides and antibiotic resistance. Tartu, 1997, 80 p.
29. **Arvo Tuvikene**. Assessment of inland water pollution using biomarker responses in fish *in vivo* and *in vitro.* Tartu, 1997, 160 p.
30. **Urmas Saarma**. Tuning ribosomal elongation cycle by mutagenesis of 23S rRNA. Tartu, 1997, 134 p.
31. **Henn Ojaveer**. Composition and dynamics of fish stocks in the gulf of Riga ecosystem. Tartu, 1997, 138 p.
32. **Lembi Lõugas**. Post-glacial development of vertebrate fauna in Estonian water bodies. Tartu, 1997, 138 p.
33. **Margus Pooga**. Cell penetrating peptide, transportan, and its predecessors, galanin-based chimeric peptides. Tartu, 1998, 110 p.
34. **Andres Saag**. Evolutionary relationships in some cetrarioid genera (Lichenized Ascomycota). Tartu, 1998, 196 p.
35. **Aivar Liiv**. Ribosomal large subunit assembly *in vivo*. Tartu, 1998, 158 p.
36. **Tatjana Oja**. Isoenzyme diversity and phylogenetic affinities among the eurasian annual bromes (*Bromus* L., Poaceae). Tartu, 1998, 92 p.
37. **Mari Moora**. The influence of arbuscular mycorrhizal (AM) symbiosis on the competition and coexistence of calcareous grassland plant species. Tartu, 1998, 78 p.
38. **Olavi Kurina**. Fungus gnats in Estonia (*Diptera: Bolitophilidae, Keroplatidae, Macroceridae, Ditomyiidae, Diadocidiidae, Mycetophilidae*). Tartu, 1998, 200 p.
39. **Andrus Tasa**. Biological leaching of shales: black shale and oil shale. Tartu, 1998, 98 p.
40. **Arnold Kristjuhan**. Studies on transcriptional activator properties of tumor suppressor protein p53. Tartu, 1998, 86 p.
41. **Sulev Ingerpuu**. Characterization of some human myeloid cell surface and nuclear differentiation antigens. Tartu, 1998, 163 p.

42. **Veljo Kisand**. Responses of planktonic bacteria to the abiotic and biotic factors in the shallow lake Võrtsjärv. Tartu, 1998, 118 p.
43. **Kadri Põldmaa**. Studies in the systematics of hypomyces and allied genera (Hypocreales, Ascomycota). Tartu, 1998, 178 p.
44. **Markus Vetemaa**. Reproduction parameters of fish as indicators in environmental monitoring. Tartu, 1998, 117 p.
45. **Heli Talvik**. Prepatent periods and species composition of different *Oesophagostomum* spp. populations in Estonia and Denmark. Tartu, 1998, 104 p.
46. **Katrin Heinsoo**. Cuticular and stomatal antechamber conductance to water vapour diffusion in *Picea abies* (L.) karst. Tartu, 1999, 133 p.
47. **Tarmo Annilo**. Studies on mammalian ribosomal protein S7. Tartu, 1998, 77 p.
48. **Indrek Ots**. Health state indicies of reproducing great tits (*Parus major*): sources of variation and connections with life-history traits. Tartu, 1999, 117 p.
49. **Juan Jose Cantero**. Plant community diversity and habitat relationships in central Argentina grasslands. Tartu, 1999, 161 p.
50. **Rein Kalamees**. Seed bank, seed rain and community regeneration in Estonian calcareous grasslands. Tartu, 1999, 107 p.
51. **Sulev Kõks**. Cholecystokinin (CCK) – induced anxiety in rats: influence of environmental stimuli and involvement of endopioid mechanisms and serotonin. Tartu, 1999, 123 p.
52. **Ebe Sild**. Impact of increasing concentrations of O_3 and CO_2 on wheat, clover and pasture. Tartu, 1999, 123 p.
53. **Ljudmilla Timofejeva**. Electron microscopical analysis of the synaptonemal complex formation in cereals. Tartu, 1999, 99 p.
54. **Andres Valkna**. Interactions of galanin receptor with ligands and G-proteins: studies with synthetic peptides. Tartu, 1999, 103 p.
55. **Taavi Virro**. Life cycles of planktonic rotifers in lake Peipsi. Tartu, 1999, 101 p.
56. **Ana Rebane**. Mammalian ribosomal protein S3a genes and intron-encoded small nucleolar RNAs U73 and U82. Tartu, 1999, 85 p.
57. **Tiina Tamm**. Cocksfoot mottle virus: the genome organisation and translational strategies. Tartu, 2000, 101 p.
58. **Reet Kurg**. Structure-function relationship of the bovine papilloma virus E2 protein. Tartu, 2000, 89 p.
59. **Toomas Kivisild**. The origins of Southern and Western Eurasian populations: an mtDNA study. Tartu, 2000, 121 p.
60. **Niilo Kaldalu**. Studies of the TOL plasmid transcription factor XylS. Tartu, 2000, 88 p.
61. **Dina Lepik**. Modulation of viral DNA replication by tumor suppressor protein p53. Tartu, 2000, 106 p.

62. **Kai Vellak**. Influence of different factors on the diversity of the bryophyte vegetation in forest and wooded meadow communities. Tartu, 2000, 122 p.
63. **Jonne Kotta**. Impact of eutrophication and biological invasionas on the structure and functions of benthic macrofauna. Tartu, 2000, 160 p.
64. **Georg Martin**. Phytobenthic communities of the Gulf of Riga and the inner sea the West-Estonian archipelago. Tartu, 2000, 139 p.
65. **Silvia Sepp**. Morphological and genetical variation of *Alchemilla L.* in Estonia. Tartu, 2000. 124 p.
66. **Jaan Liira**. On the determinants of structure and diversity in herbaceous plant communities. Tartu, 2000, 96 p.
67. **Priit Zingel**. The role of planktonic ciliates in lake ecosystems. Tartu, 2001, 111 p.
68. **Tiit Teder**. Direct and indirect effects in Host-parasitoid interactions: ecological and evolutionary consequences. Tartu, 2001, 122 p.
69. **Hannes Kollist**. Leaf apoplastic ascorbate as ozone scavenger and its transport across the plasma membrane. Tartu, 2001, 80 p.
70. **Reet Marits**. Role of two-component regulator system PehR-PehS and extracellular protease PrtW in virulence of *Erwinia Carotovora* subsp. *Carotovora*. Tartu, 2001, 112 p.
71. **Vallo Tilgar**. Effect of calcium supplementation on reproductive performance of the pied flycatcher *Ficedula hypoleuca* and the great tit *Parus major,* breeding in Nothern temperate forests. Tartu, 2002, 126 p.
72. **Rita Hõrak**. Regulation of transposition of transposon Tn*4652* in *Pseudomonas putida*. Tartu, 2002, 108 p.
73. **Liina Eek-Piirsoo**. The effect of fertilization, mowing and additional illumination on the structure of a species-rich grassland community. Tartu, 2002, 74 p.
74. **Krõõt Aasamaa**. Shoot hydraulic conductance and stomatal conductance of six temperate deciduous tree species. Tartu, 2002, 110 p.
75. **Nele Ingerpuu**. Bryophyte diversity and vascular plants. Tartu, 2002, 112 p.
76. **Neeme Tõnisson**. Mutation detection by primer extension on oligonucleotide microarrays. Tartu, 2002, 124 p.
77. **Margus Pensa**. Variation in needle retention of Scots pine in relation to leaf morphology, nitrogen conservation and tree age. Tartu, 2003, 110 p.
78. **Asko Lõhmus**. Habitat preferences and quality for birds of prey: from principles to applications. Tartu, 2003, 168 p.
79. **Viljar Jaks**. p53 – a switch in cellular circuit. Tartu, 2003, 160 p.
80. **Jaana Männik**. Characterization and genetic studies of four ATP-binding cassette (ABC) transporters. Tartu, 2003, 140 p.
81. **Marek Sammul**. Competition and coexistence of clonal plants in relation to productivity. Tartu, 2003, 159 p
82. **Ivar Ilves**. Virus-cell interactions in the replication cycle of bovine papillomavirus type 1. Tartu, 2003, 89 p.

83. **Andres Männik**. Design and characterization of a novel vector system based on the stable replicator of bovine papillomavirus type 1. Tartu, 2003, 109 p.
84. **Ivika Ostonen**. Fine root structure, dynamics and proportion in net primary production of Norway spruce forest ecosystem in relation to site conditions. Tartu, 2003, 158 p.
85. **Gudrun Veldre**. Somatic status of 12–15-year-old Tartu schoolchildren. Tartu, 2003, 199 p.
86. **Ülo Väli**. The greater spotted eagle *Aquila clanga* and the lesser spotted eagle *A. pomarina*: taxonomy, phylogeography and ecology. Tartu, 2004, 159 p.
87. **Aare Abroi**. The determinants for the native activities of the bovine papillomavirus type 1 E2 protein are separable. Tartu, 2004, 135 p.
88. **Tiina Kahre**. Cystic fibrosis in Estonia. Tartu, 2004, 116 p.
89. **Helen Orav-Kotta**. Habitat choice and feeding activity of benthic suspension feeders and mesograzers in the northern Baltic Sea. Tartu, 2004, 117 p.
90. **Maarja Öpik**. Diversity of arbuscular mycorrhizal fungi in the roots of perennial plants and their effect on plant performance. Tartu, 2004, 175 p.
91. **Kadri Tali**. Species structure of *Neotinea ustulata*. Tartu, 2004, 109 p.
92. **Kristiina Tambets**. Towards the understanding of post-glacial spread of human mitochondrial DNA haplogroups in Europe and beyond: a phylogeographic approach. Tartu, 2004, 163 p.
93. **Arvi Jõers**. Regulation of p53-dependent transcription. Tartu, 2004, 103 p.
94. **Lilian Kadaja**. Studies on modulation of the activity of tumor suppressor protein p53. Tartu, 2004, 103 p.
95. **Jaak Truu**. Oil shale industry wastewater: impact on river microbial community and possibilities for bioremediation. Tartu, 2004, 128 p.
96. **Maire Peters**. Natural horizontal transfer of the *pheBA* operon. Tartu, 2004, 105 p.
97. **Ülo Maiväli**. Studies on the structure-function relationship of the bacterial ribosome. Tartu, 2004, 130 p.
98. **Merit Otsus**. Plant community regeneration and species diversity in dry calcareous grasslands. Tartu, 2004, 103 p.
99. **Mikk Heidemaa**. Systematic studies on sawflies of the genera *Dolerus, Empria,* and *Caliroa* (Hymenoptera: Tenthredinidae). Tartu, 2004, 167 p.
100. **Ilmar Tõnno**. The impact of nitrogen and phosphorus concentration and N/P ratio on cyanobacterial dominance and N_2 fixation in some Estonian lakes. Tartu, 2004, 111 p.
101. **Lauri Saks**. Immune function, parasites, and carotenoid-based ornaments in greenfinches. Tartu, 2004, 144 p.
102. **Siiri Rootsi**. Human Y-chromosomal variation in European populations. Tartu, 2004, 142 p.
103. **Eve Vedler**. Structure of the 2,4-dichloro-phenoxyacetic acid-degradative plasmid pEST4011. Tartu, 2005. 106 p.

104. **Andres Tover**. Regulation of transcription of the phenol degradation *pheBA* operon in *Pseudomonas putida*. Tartu, 2005, 126 p.
105. **Helen Udras**. Hexose kinases and glucose transport in the yeast *Hansenula polymorpha*. Tartu, 2005, 100 p.
106. **Ave Suija**. Lichens and lichenicolous fungi in Estonia: diversity, distribution patterns, taxonomy. Tartu, 2005, 162 p.
107. **Piret Lõhmus**. Forest lichens and their substrata in Estonia. Tartu, 2005, 162 p.
108. **Inga Lips**. Abiotic factors controlling the cyanobacterial bloom occurrence in the Gulf of Finland. Tartu, 2005, 156 p.
109. **Kaasik, Krista**. Circadian clock genes in mammalian clockwork, metabolism and behaviour. Tartu, 2005, 121 p.
110. **Juhan Javoiš**. The effects of experience on host acceptance in ovipositing moths. Tartu, 2005, 112 p.
111. **Tiina Sedman**. Characterization of the yeast *Saccharomyces cerevisiae* mitochondrial DNA helicase Hmi1. Tartu, 2005, 103 p.
112. **Ruth Aguraiuja**. Hawaiian endemic fern lineage *Diellia* (Aspleniaceae): distribution, population structure and ecology. Tartu, 2005, 112 p.
113. **Riho Teras**. Regulation of transcription from the fusion promoters generated by transposition of Tn*4652* into the upstream region of *pheBA* operon in *Pseudomonas putida*. Tartu, 2005, 106 p.
114. **Mait Metspalu**. Through the course of prehistory in india: tracing the mtDNA trail. Tartu, 2005, 138 p.
115. **Elin Lõhmussaar**. The comparative patterns of linkage disequilibrium in European populations and its implication for genetic association studies. Tartu, 2006, 124 p.
116. **Priit Kupper**. Hydraulic and environmental limitations to leaf water relations in trees with respect to canopy position. Tartu, 2006, 126 p.
117. **Heili Ilves**. Stress-induced transposition of Tn*4652* in *Pseudomonas Putida.* Tartu, 2006, 120 p.
118. **Silja Kuusk**. Biochemical properties of Hmi1p, a DNA helicase from *Saccharomyces cerevisiae* mitochondria. Tartu, 2006, 126 p.
119. **Kersti Püssa**. Forest edges on medium resolution landsat thematic mapper satellite images. Tartu, 2006, 90 p.
120. **Lea Tummeleht**. Physiological condition and immune function in great tits (*Parus major* l.): Sources of variation and trade-offs in relation to growth. Tartu, 2006, 94 p.
121. **Toomas Esperk**. Larval instar as a key element of insect growth schedules. Tartu, 2006, 186 p.
122. **Harri Valdmann**. Lynx (*Lynx lynx*) and wolf (*Canis lupus*) in the Baltic region: Diets, helminth parasites and genetic variation. Tartu, 2006. 102 p.
123. **Priit Jõers**. Studies of the mitochondrial helicase Hmi1p in *Candida albicans* and *Saccharomyces cerevisia*. Tartu, 2006. 113 p.
124. **Kersti Lilleväli**. Gata3 and Gata2 in inner ear development. Tartu, 2007, 123 p.

125. **Kai Rünk**. Comparative ecology of three fern species: *Dryopteris carthusiana* (Vill.) H.P. Fuchs, *D. expansa* (C. Presl) Fraser-Jenkins & Jermy and *D. dilatata* (Hoffm.) A. Gray (Dryopteridaceae). Tartu, 2007, 143 p.
126. **Aveliina Helm**. Formation and persistence of dry grassland diversity: role of human history and landscape structure. Tartu, 2007, 89 p.
127. **Leho Tedersoo**. Ectomycorrhizal fungi: diversity and community structure in Estonia, Seychelles and Australia. Tartu, 2007, 233 p.
128. **Marko Mägi**. The habitat-related variation of reproductive performance of great tits in a deciduous-coniferous forest mosaic: looking for causes and consequences. Tartu, 2007, 135 p.
129. **Valeria Lulla**. Replication strategies and applications of Semliki Forest virus. Tartu, 2007, 109 p.
130. **Ülle Reier**. Estonian threatened vascular plant species: causes of rarity and conservation. Tartu, 2007, 79 p.
131. **Inga Jüriado**. Diversity of lichen species in Estonia: influence of regional and local factors. Tartu, 2007, 171 p.
132. **Tatjana Krama**. Mobbing behaviour in birds: costs and reciprocity based cooperation. Tartu, 2007, 112 p.
133. **Signe Saumaa**. The role of DNA mismatch repair and oxidative DNA damage defense systems in avoidance of stationary phase mutations in *Pseudomonas putida.* Tartu, 2007, 172 p.
134. **Reedik Mägi**. The linkage disequilibrium and the selection of genetic markers for association studies in european populations. Tartu, 2007, 96 p.
135. **Priit Kilgas**. Blood parameters as indicators of physiological condition and skeletal development in great tits (*Parus major*): natural variation and application in the reproductive ecology of birds. Tartu, 2007, 129 p.
136. **Anu Albert**. The role of water salinity in structuring eastern Baltic coastal fish communities. Tartu, 2007, 95 p.
137. **Kärt Padari**. Protein transduction mechanisms of transportans. Tartu, 2008, 128 p.
138. **Siiri-Lii Sandre**. Selective forces on larval colouration in a moth. Tartu, 2008, 125 p.
139. **Ülle Jõgar**. Conservation and restoration of semi-natural floodplain meadows and their rare plant species. Tartu, 2008, 99 p.
140. **Lauri Laanisto**. Macroecological approach in vegetation science: generality of ecological relationships at the global scale. Tartu, 2008, 133 p.
141. **Reidar Andreson**. Methods and software for predicting PCR failure rate in large genomes. Tartu, 2008, 105 p.
142. **Birgot Paavel**. Bio-optical properties of turbid lakes. Tartu, 2008, 175 p.
143. **Kaire Torn**. Distribution and ecology of charophytes in the Baltic Sea. Tartu, 2008, 98 p.
144. **Vladimir Vimberg**. Peptide mediated macrolide resistance. Tartu, 2008, 190 p.
145. **Daima Örd**. Studies on the stress-inducible pseudokinase TRB3, a novel inhibitor of transcription factor ATF4. Tartu, 2008, 108 p.

146. **Lauri Saag**. Taxonomic and ecologic problems in the genus *Lepraria* (*Stereocaulaceae*, lichenised *Ascomycota*). Tartu, 2008, 175 p.
147. **Ulvi Karu**. Antioxidant protection, carotenoids and coccidians in greenfinches – assessment of the costs of immune activation and mechanisms of parasite resistance in a passerine with carotenoid-based ornaments. Tartu, 2008, 124 p.
148. **Jaanus Remm**. Tree-cavities in forests: density, characteristics and occupancy by animals. Tartu, 2008, 128 p.
149. **Epp Moks**. Tapeworm parasites *Echinococcus multilocularis* and *E. granulosus* in Estonia: phylogenetic relationships and occurrence in wild carnivores and ungulates. Tartu, 2008, 82 p.
150. **Eve Eensalu**. Acclimation of stomatal structure and function in tree canopy: effect of light and CO_2 concentration. Tartu, 2008, 108 p.
151. **Janne Pullat**. Design, functionlization and application of an *in situ* synthesized oligonucleotide microarray. Tartu, 2008, 108 p.
152. **Marta Putrinš**. Responses of *Pseudomonas putida* to phenol-induced metabolic and stress signals. Tartu, 2008, 142 p.
153. **Marina Semtšenko**. Plant root behaviour: responses to neighbours and physical obstructions. Tartu, 2008, 106 p.
154. **Marge Starast**. Influence of cultivation techniques on productivity and fruit quality of some *Vaccinium* and *Rubus* taxa. Tartu, 2008, 154 p.
155. **Age Tats**. Sequence motifs influencing the efficiency of translation. Tartu, 2009, 104 p.
156. **Radi Tegova**. The role of specialized DNA polymerases in mutagenesis in *Pseudomonas putida.* Tartu, 2009, 124 p.
157. **Tsipe Aavik**. Plant species richness, composition and functional trait pattern in agricultural landscapes – the role of land use intensity and landscape structure. Tartu, 2009, 112 p.
158. **Kaja Kiiver**. Semliki forest virus based vectors and cell lines for studying the replication and interactions of alphaviruses and hepaciviruses. Tartu, 2009, 104 p.
159. **Meelis Kadaja**. Papillomavirus Replication Machinery Induces Genomic Instability in its Host Cell. Tartu, 2009, 126 p.
160. **Pille Hallast**. Human and chimpanzee Luteinizing hormone/Chorionic Gonadotropin beta (*LHB/CGB*) gene clusters: diversity and divergence of young duplicated genes. Tartu, 2009, 168 p.
161. **Ain Vellak**. Spatial and temporal aspects of plant species conservation. Tartu, 2009, 86 p.
162. **Triinu Remmel**. Body size evolution in insects with different colouration strategies: the role of predation risk. Tartu, 2009, 168 p.
163. **Jaana Salujõe**. Zooplankton as the indicator of ecological quality and fish predation in lake ecosystems. Tartu, 2009, 129 p.
164. **Ele Vahtmäe**. Mapping benthic habitat with remote sensing in optically complex coastal environments. Tartu, 2009, 109 p.

165. **Liisa Metsamaa**. Model-based assessment to improve the use of remote sensing in recognition and quantitative mapping of cyanobacteria. Tartu, 2009, 114 p.
166. **Pille Säälik**. The role of endocytosis in the protein transduction by cell-penetrating peptides. Tartu, 2009, 155 p.
167. **Lauri Peil**. Ribosome assembly factors in *Escherichia coli.* Tartu, 2009, 147 p.
168. **Lea Hallik**. Generality and specificity in light harvesting, carbon gain capacity and shade tolerance among plant functional groups. Tartu, 2009, 99 p.
169. **Mariliis Tark**. Mutagenic potential of DNA damage repair and tolerance mechanisms under starvation stress. Tartu, 2009, 191 p.
170. **Riinu Rannap**. Impacts of habitat loss and restoration on amphibian populations. Tartu, 2009, 117 p.
171. **Maarja Adojaan**. Molecular variation of HIV-1 and the use of this knowledge in vaccine development. Tartu, 2009, 95 p.
172. **Signe Altmäe**. Genomics and transcriptomics of human induced ovarian folliculogenesis. Tartu, 2010, 179 p.
173. **Triin Suvi**. Mycorrhizal fungi of native and introduced trees in the Seychelles Islands. Tartu, 2010, 107 p.
174. **Velda Lauringson**. Role of suspension feeding in a brackish-water coastal sea. Tartu, 2010, 123 p.
175. **Eero Talts**. Photosynthetic cyclic electron transport – measurement and variably proton-coupled mechanism. Tartu, 2010, 121 p.
176. **Mari Nelis**. Genetic structure of the Estonian population and genetic distance from other populations of European descent. Tartu, 2010, 97 p.
177. **Kaarel Krjutškov**. Arrayed Primer Extension-2 as a multiplex PCR-based method for nucleic acid variation analysis: method and applications. Tartu, 2010, 129 p.
178. **Egle Köster**. Morphological and genetical variation within species complexes: *Anthyllis vulneraria* s. l. and *Alchemilla vulgaris* (coll.). Tartu, 2010, 101 p.
179. **Erki Õunap**. Systematic studies on the subfamily Sterrhinae (Lepidoptera: Geometridae). Tartu, 2010, 111 p.
180. **Merike Jõesaar**. Diversity of key catabolic genes at degradation of phenol and *p*-cresol in pseudomonads. Tartu, 2010, 125 p.
181. **Kristjan Herkül**. Effects of physical disturbance and habitat-modifying species on sediment properties and benthic communities in the northern Baltic Sea. Tartu, 2010, 123 p.
182. **Arto Pulk**. Studies on bacterial ribosomes by chemical modification approaches. Tartu, 2010, 161 p.
183. **Maria Põllupüü**. Ecological relations of cladocerans in a brackish-water ecosystem. Tartu, 2010, 126 p.
184. **Toomas Silla**. Study of the segregation mechanism of the Bovine Papillomavirus Type 1. Tartu, 2010, 188 p.

185. **Gyaneshwer Chaubey**. The demographic history of India: A perspective based on genetic evidence. Tartu, 2010, 184 p.
186. **Katrin Kepp**. Genes involved in cardiovascular traits: detection of genetic variation in Estonian and Czech populations. Tartu, 2010, 164 p.
187. **Virve Sõber**. The role of biotic interactions in plant reproductive performance. Tartu, 2010, 92 p.
188. **Kersti Kangro**. The response of phytoplankton community to the changes in nutrient loading. Tartu, 2010, 144 p.
189. **Joachim M. Gerhold**. Replication and Recombination of mitochondrial DNA in Yeast. Tartu, 2010, 120 p.
190. **Helen Tammert**. Ecological role of physiological and phylogenetic diversity in aquatic bacterial communities. Tartu, 2010, 140 p.
191. **Elle Rajandu**. Factors determining plant and lichen species diversity and composition in Estonian *Calamagrostis* and *Hepatica* site type forests. Tartu, 2010, 123 p.
192. **Paula Ann Kivistik**. ColR-ColS signalling system and transposition of Tn*4652* in the adaptation of *Pseudomonas putida.* Tartu, 2010, 118 p.
193. **Siim Sõber**. Blood pressure genetics: from candidate genes to genome-wide association studies. Tartu, 2011, 120 p.
194. **Kalle Kipper**. Studies on the role of helix 69 of 23S rRNA in the factor-dependent stages of translation initiation, elongation, and termination. Tartu, 2011, 178 p.
195. **Triinu Siibak**. Effect of antibiotics on ribosome assembly is indirect. Tartu, 2011, 134 p.
196. **Tambet Tõnissoo**. Identification and molecular analysis of the role of guanine nucleotide exchange factor RIC-8 in mouse development and neural function. Tartu, 2011, 110 p.
197. **Helin Räägel**. Multiple faces of cell-penetrating peptides – their intracellular trafficking, stability and endosomal escape during protein transduction. Tartu, 2011, 161 p.
198. **Andres Jaanus**. Phytoplankton in Estonian coastal waters – variability, trends and response to environmental pressures. Tartu, 2011, 157 p.
199. **Tiit Nikopensius**. Genetic predisposition to nonsyndromic orofacial clefts. Tartu, 2011, 152 p.
200. **Signe Värv**. Studies on the mechanisms of RNA polymerase II-dependent transcription elongation. Tartu, 2011, 108 p.
201. **Kristjan Välk**. Gene expression profiling and genome-wide association studies of non-small cell lung cancer. Tartu, 2011, 98 p.
202. **Arno Põllumäe**. Spatio-temporal patterns of native and invasive zooplankton species under changing climate and eutrophication conditions. Tartu, 2011, 153 p.
203. **Egle Tammeleht**. Brown bear (*Ursus arctos*) population structure, demographic processes and variations in diet in northern Eurasia. Tartu, 2011, 143 p.

205. **Teele Jairus**. Species composition and host preference among ectomycorrhizal fungi in Australian and African ecosystems. Tartu, 2011, 106 p.
206. **Kessy Abarenkov**. PlutoF – cloud database and computing services supporting biological research. Tartu, 2011, 125 p.
207. **Marina Grigorova**. Fine-scale genetic variation of follicle-stimulating hormone beta-subunit coding gene (*FSHB*) and its association with reproductive health. Tartu, 2011, 184 p.
208. **Anu Tiitsaar**. The effects of predation risk and habitat history on butterfly communities. Tartu, 2011, 97 p.
209. **Elin Sild**. Oxidative defences in immunoecological context: validation and application of assays for nitric oxide production and oxidative burst in a wild passerine. Tartu, 2011, 105 p.
210. **Irja Saar**. The taxonomy and phylogeny of the genera *Cystoderma* and *Cystodermella* (Agaricales, Fungi). Tartu, 2012, 167 p.
211. **Pauli Saag**. Natural variation in plumage bacterial assemblages in two wild breeding passerines. Tartu, 2012, 113 p.
212. **Aleksei Lulla**. Alphaviral nonstructural protease and its polyprotein substrate: arrangements for the perfect marriage. Tartu, 2012, 143 p.
213. **Mari Järve**. Different genetic perspectives on human history in Europe and the Caucasus: the stories told by uniparental and autosomal markers. Tartu, 2012, 119 p.
214. **Ott Scheler**. The application of tmRNA as a marker molecule in bacterial diagnostics using microarray and biosensor technology. Tartu, 2012, 93 p.
215. **Anna Balikova**. Studies on the functions of tumor-associated mucin-like leukosialin (CD43) in human cancer cells. Tartu, 2012, 129 p.
216. **Triinu Kõressaar**. Improvement of PCR primer design for detection of prokaryotic species. Tartu, 2012, 83 p.
217. **Tuul Sepp**. Hematological health state indices of greenfinches: sources of individual variation and responses to immune system manipulation. Tartu, 2012, 117 p.
218. **Rya Ero**. Modifier view of the bacterial ribosome. Tartu, 2012, 146 p.
219. **Mohammad Bahram**. Biogeography of ectomycorrhizal fungi across different spatial scales. Tartu, 2012, 165 p.
220. **Annely Lorents**. Overcoming the plasma membrane barrier: uptake of amphipathic cell-penetrating peptides induces influx of calcium ions and downstream responses. Tartu, 2012, 113 p.
221. **Katrin Männik**. Exploring the genomics of cognitive impairment: whole-genome SNP genotyping experience in Estonian patients and general population. Tartu, 2012, 171 p.
222. **Marko Prous**. Taxonomy and phylogeny of the sawfly genus *Empria* (Hymenoptera, Tenthredinidae). Tartu, 2012, 192 p.
223. **Triinu Visnapuu**. Levansucrases encoded in the genome of *Pseudomonas syringae* pv. tomato DC3000: heterologous expression, biochemical characterization, mutational analysis and spectrum of polymerization products. Tartu, 2012, 160 p.

224. **Nele Tamberg**. Studies on Semliki Forest virus replication and pathogenesis. Tartu, 2012, 109 p.
225. **Tõnu Esko**. Novel applications of SNP array data in the analysis of the genetic structure of Europeans and in genetic association studies. Tartu, 2012, 149 p.
226. **Timo Arula**. Ecology of early life-history stages of herring *Clupea harengus membras* in the northeastern Baltic Sea. Tartu, 2012, 143 p.
227. **Inga Hiiesalu**. Belowground plant diversity and coexistence patterns in grassland ecosystems. Tartu, 2012, 130 p.
228. **Kadri Koorem**. The influence of abiotic and biotic factors on small-scale plant community patterns and regeneration in boreonemoral forest. Tartu, 2012, 114 p.
229. **Liis Andresen**. Regulation of virulence in plant-pathogenic pectobacteria. Tartu, 2012, 122 p.
230. **Kaupo Kohv**. The direct and indirect effects of management on boreal forest structure and field layer vegetation. Tartu, 2012, 124 p.
231. **Mart Jüssi**. Living on an edge: landlocked seals in changing climate. Tartu, 2012, 114 p.
232. **Riina Klais**. Phytoplankton trends in the Baltic Sea. Tartu, 2012, 136 p.
233. **Rauno Veeroja**. Effects of winter weather, population density and timing of reproduction on life-history traits and population dynamics of moose (*Alces alces*) in Estonia. Tartu, 2012, 92 p.
234. **Marju Keis**. Brown bear (*Ursus arctos*) phylogeography in northern Eurasia. Tartu, 2013, 142 p.
235. **Sergei Põlme**. Biogeography and ecology of *alnus*- associated ectomycorrhizal fungi – from regional to global scale. Tartu, 2013, 90 p.
236. **Liis Usküla**. Placental gene expression in normal and complicated pregnancy. Tartu, 2013, 173 p.
237. **Marko Lõoke**. Studies on DNA replication initiation in *Saccharomyces cerevisiae.* Tartu, 2013, 112 p.
238. **Anne Aan**. Light- and nitrogen-use and biomass allocation along productivity gradients in multilayer plant communities. Tartu, 2013, 127 p.
239. **Heidi Tamm**. Comprehending phylogenetic diversity – case studies in three groups of ascomycetes. Tartu, 2013, 136 p.
240. **Liina Kangur**. High-Pressure Spectroscopy Study of Chromophore-Binding Hydrogen Bonds in Light-Harvesting Complexes of Photosynthetic Bacteria. Tartu, 2013, 150 p.
241. **Margus Leppik**. Substrate specificity of the multisite specific pseudouridine synthase RluD. Tartu, 2013, 111 p.
242. **Lauris Kaplinski**. The application of oligonucleotide hybridization model for PCR and microarray optimization. Tartu, 2013, 103 p.
243. **Merli Pärnoja**. Patterns of macrophyte distribution and productivity in coastal ecosystems: effect of abiotic and biotic forcing. Tartu, 2013, 155 p.
244. **Tõnu Margus**. Distribution and phylogeny of the bacterial translational GTPases and the Mqsr/YgiT regulatory system. Tartu, 2013, 126 p.

245. **Pille Mänd**. Light use capacity and carbon and nitrogen budget of plants: remote assessment and physiological determinants. Tartu, 2013, 128 p.
246. **Mario Plaas**. Animal model of Wolfram Syndrome in mice: behavioural, biochemical and psychopharmacological characterization. Tartu, 2013, 144 p.
247. **Georgi Hudjašov**. Maps of mitochondrial DNA, Y-chromosome and tyrosinase variation in Eurasian and Oceanian populations. Tartu, 2013, 115 p.
248. **Mari Lepik**. Plasticity to light in herbaceous plants and its importance for community structure and diversity. Tartu, 2013, 102 p.
249. **Ede Leppik**. Diversity of lichens in semi-natural habitats of Estonia. Tartu, 2013, 151 p.
250. **Ülle Saks**. Arbuscular mycorrhizal fungal diversity patterns in boreo-nemoral forest ecosystems. Tartu, 2013, 151 p.
251. **Eneli Oitmaa**. Development of arrayed primer extension microarray assays for molecular diagnostic applications. Tartu, 2013, 147 p.
252. **Jekaterina Jutkina**. The horizontal gene pool for aromatics degradation: bacterial catabolic plasmids of the Baltic Sea aquatic system. Tartu, 2013, 121 p.
253. **Helen Vellau**. Reaction norms for size and age at maturity in insects: rules and exceptions. Tartu, 2014, 132 p.
254. **Randel Kreitsberg**. Using biomarkers in assessment of environmental contamination in fish – new perspectives. Tartu, 2014, 107 p.
255. **Krista Takkis**. Changes in plant species richness and population performance in response to habitat loss and fragmentation.Tartu, 2014, 141 p.
256. **Liina Nagirnaja**. Global and fine-scale genetic determinants of recurrent pregnancy loss. Tartu, 2014, 211 p.
257. **Triin Triisberg**. Factors influencing the re-vegetation of abandoned extracted peatlands in Estonia. Tartu, 2014, 133 p.
258. **Villu Soon**. A phylogenetic revision of the *Chrysis ignita* species group (Hymenoptera: Chrysididae) with emphasis on the northern European fauna. Tartu, 2014, 211 p.
259. **Andrei Nikonov**. RNA-Dependent RNA Polymerase Activity as a Basis for the Detection of Positive-Strand RNA Viruses by Vertebrate Host Cells. Tartu, 2014, 207 p.
260. **Eele Õunapuu-Pikas**. Spatio-temporal variability of leaf hydraulic conductance in woody plants: ecophysiological consequences. Tartu, 2014, 135 p.
261. **Marju Männiste**. Physiological ecology of greenfinches: information content of feathers in relation to immune function and behavior. Tartu, 2014, 121 p.
262. **Katre Kets**. Effects of elevated concentrations of CO_2 and O_3 on leaf photosynthetic parameters in *Populus tremuloides*: diurnal, seasonal and interannual patterns. Tartu, 2014, 115 p.

263. **Külli Lokko**. Seasonal and spatial variability of zoopsammon communities in relation to environmental parameters. Tartu, 2014, 129 p.
264. **Olga Žilina**. Chromosomal microarray analysis as diagnostic tool: Estonian experience. Tartu, 2014, 152 p.
265. **Kertu Lõhmus**. Colonisation ecology of forest-dwelling vascular plants and the conservation value of rural manor parks. Tartu, 2014, 111 p.
266. **Anu Aun**. Mitochondria as integral modulators of cellular signaling. Tartu, 2014, 167 p.
267. **Chandana Basu Mallick**. Genetics of adaptive traits and gender-specific demographic processes in South Asian populations. Tartu, 2014, 160 p.
268. **Riin Tamme**. The relationship between small-scale environmental heterogeneity and plant species diversity. Tartu, 2014, 130 p.
269. **Liina Remm**. Impacts of forest drainage on biodiversity and habitat quality: implications for sustainable management and conservation. Tartu, 2015, 126 p.
270. **Tiina Talve**. Genetic diversity and taxonomy within the genus *Rhinanthus.* Tartu, 2015, 106 p.
271. **Mehis Rohtla**. Otolith sclerochronological studies on migrations, spawning habitat preferences and age of freshwater fishes inhabiting the Baltic Sea. Tartu, 2015, 137 p.
272. **Alexey Reshchikov**. The world fauna of the genus *Lathrolestes* (Hymenoptera, Ichneumonidae). Tartu, 2015, 247 p.
273. **Martin Pook**. Studies on artificial and extracellular matrix protein-rich surfaces as regulators of cell growth and differentiation. Tartu, 2015, 142 p.
274. **Mai Kukumägi**. Factors affecting soil respiration and its components in silver birch and Norway spruce stands. Tartu, 2015, 155 p.
275. **Helen Karu**. Development of ecosystems under human activity in the North-East Estonian industrial region: forests on post-mining sites and bogs. Tartu, 2015, 152 p.
276. **Hedi Peterson**. Exploiting high-throughput data for establishing relationships between genes. Tartu, 2015, 186 p.
277. **Priit Adler**. Analysis and visualisation of large scale microarray data, Tartu, 2015, 126 p.
278. **Aigar Niglas**. Effects of environmental factors on gas exchange in deciduous trees: focus on photosynthetic water-use efficiency. Tartu, 2015, 152 p.
279. **Silja Laht**. Classification and identification of conopeptides using profile hidden Markov models and position-specific scoring matrices. Tartu, 2015, 100 p.
280. **Martin Kesler**. Biological characteristics and restoration of Atlantic salmon *Salmo salar* populations in the Rivers of Northern Estonia. Tartu, 2015, 97 p.
281. **Pratyush Kumar Das**. Biochemical perspective on alphaviral nonstructural protein 2: a tale from multiple domains to enzymatic profiling. Tartu, 2015, 205 p

282. **Priit Palta**. Computational methods for DNA copy number detection. Tartu, 2015, 130 p.
283. **Julia Sidorenko**. Combating DNA damage and maintenance of genome integrity in pseudomonads. Tartu, 2015, 174 p.
284. **Anastasiia Kovtun-Kante**. Charophytes of Estonian inland and coastal waters: distribution and environmental preferences. Tartu, 2015, 97 p.
285. **Ly Lindman**. The ecology of protected butterfly species in Estonia. Tartu, 2015, 171 p.
286. **Jaanis Lodjak**. Association of Insulin-like Growth Factor I and Corticosterone with Nestling Growth and Fledging Success in Wild Passerines. Tartu, 2016, 113 p.
287. **Ann Kraut**. Conservation of Wood-Inhabiting Biodiversity – Semi-Natural Forests as an Opportunity. Tartu, 2016, 141 p.
288. **Tiit Örd.** Functions and regulation of the mammalian pseudokinase TRIB3. Tartu, 2016, 182. p.
289. **Kairi Käiro.** Biological Quality According to Macroinvertebrates in Streams of Estonia (Baltic Ecoregion of Europe): Effects of Human-induced Hydromorphological Changes. Tartu, 2016, 126 p.
290. **Leidi Laurimaa**. *Echinococcus multilocularis* and other zoonotic parasites in Estonian canids. Tartu, 2016, 144 p.
291. **Helerin Margus.** Characterization of cell-penetrating peptide/nucleic acid nanocomplexes and their cell-entry mechanisms. Tartu, 2016, 173 p.
292. **Kadri Runnel**. Fungal targets and tools for forest conservation. Tartu, 2016, 157 p.