Bioinformatics analysis of HPV associated host microRNA functions and identification of viral microRNA

Kui Qian

Institute of Biotechnology
and
Department of Biosciences
Faculty of Biological and Environmental Sciences
and
Viikki Graduate School in Biosciences
University of Helsinki

Academic Dissertation in Genetics

To be presented for public examination with the permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki in SALI 4 of Viikki Infocenter, Viikinkaari 11, Helsinki, on October 25th, 2013 at 12 noon.

Supervisors: Docent Petri Auvinen
Institute of Biotechnology
University of Helsinki
Helsinki, Finland

Docent Dario Greco
Finnish Institute of Occupational Health
Helsinki, Finland

Reviewers: Docent Janna Saarela
Institute for Molecular Medicine Finland
University of Helsinki
Helsinki, Finland

Professor Veijo Hukkanen
Department of Virology
University of Turku
Turku, Finland

Opponent: Docente Rosa Visone
Faculty of Medicine and Surgery
"G. d'Annunzio" University of Chieti-Pescara
Chieti and Pescara, Italy

Custos: Professor Tapio Palva
Department of Biosciences
Faculty of Biological and Environmental Sciences
University of Helsinki
Helsinki, Finland

ISSN 1799-7372
ISBN 978-952-10-9282-4 (paperback)
ISBN 978-952-10-9283-1 (PDF, http://ethesis.helsinki.fi)

E-mail: kui.qian@helsinki.fi

Printed: Helsinki University Press, Helsinki 2013

To my parents

TABLE OF CONTENTS
LIST OF ORIGINAL ARTICLES
ABSTRACT
ABBREVIATIONS

1. Introduction 9
1.1 Human Papillomavirus (HPV) 9
1.1.1 HPV genome 9
1.1.2 HPV oncogenes 9
1.1.3 Biology of cervical cancer 11
1.1.4 HPVs and cervical cancer 12
1.2 microRNA biogenesis and functions 12
1.2.1 Human microRNA biogenesis 13
1.2.2 Human microRNA functions 14
1.2.3 miRNAs in carcinogenesis and in human cancers 14
1.2.4 Cervical cancer and cellular miRNA 15
1.2.5 Viral microRNA biogenesis 15
1.2.6 Viral microRNA functions 16
1.3 Gene expression profiling technology 17
1.3.1 Microarray 17
1.3.1.1 mRNA and miRNA microarrays 18
1.3.2 Next Generation Sequencing 19
1.3.2.1 miRNA sequencing 20
2. Aims of the study 22
3. Methods 23
3.1 Microarray analysis (I) 23
3.2 Functional annotation analysis of gene lists (I, III) 23
3.3 Prediction of cellular microRNA targets (I) 24
3.4 miRNA sequencing data analysis workflow design (II) 24
3.5 miRNA sequencing by SOLiD platform (II) 25
3.6 Mapping and preprocessing the SOLiD sequencing data (III) 25
3.7 Prediction of candidate viral miRNAs (III) 25
3.8 Prediction of novel viral miRNA targets (III) 25
4. Result 26
4.1 E5 effects on human miRNA expression 26
4.2 Integration of differentially expressed mRNAs and miRNAs 26
4.3 miRSeqNovel workflow and its performance on public data 29
4.4 SOLiD sequencing of HPV samples 30
4.5 Prediction of novel viral miRNA candidates 30
4.6 Prediction of viral miRNA targets 32
5. Discussion 33
6. Conclusions 40
7. Acknowledgements 41
8. References 43

LIST OF ORIGINAL ARTICLES

The thesis is based on the following articles, which are referred to in the text by their Roman numerals.

I. Greco D*, Kivi N*, **Qian K**, Leivonen SK, Auvinen P, Auvinen E. Human papillomavirus 16 E5 modulates the expression of host miRNA. PLoS One. 2011;6(7):e21646.
* equal contribution

II. **Qian K**, Auvinen E, Greco D, Auvinen P. miRSeqNovel: an R based workflow for analyzing miRNA sequencing data. Mol Cell Probes. 2012 Oct;26(5):208-11.

III. **Qian K***, Pietilä T*, Rönty M, Michon F, Frilander MJ, Ritari J, Tarkkanen J, Paulín L, Auvinen P, Auvinen E. Identification and Validation of Human Papillomavirus Encoded microRNAs. PLoS One. 2013 Jul 30;8(7):e70202.
* equal contribution

Publication I was also used in the dissertation of PhD. Niina Kivi.

AUTHOR'S CONTRIBUTION TO EACH PUBLICATION

I. KQ has analyzed the microarray data, participated in the design of the experiments, and written the manuscript.

II. KQ has designed the software, collected the data, carried out all the analysis, and written the manuscript.

III. KQ has designed the experiment, carried out sequencing data analysis, and written the manuscript.

ABSTRACT

Human papillomaviruses (HPVs) form a large family of double stranded DNA (dsDNA) viruses, some types of which are the major causes of cervical cancer. HPV 16 is widely distributed and the most common high-risk HPV type and approximately half of the cervical cancers are associated with HPV type 16. Of the three HPV 16 encoded oncogenes, the function of E5 in regulating viral replication and pathogenesis is less well understood than E6 and E7. The microRNAs (miRNAs) are important small noncoding RNA molecules that regulate wide range of cellular functions. Some dsDNA viruses, such as SV40 and human polyomaviruses, have functional viral miRNAs. The functional and molecular similarities among dsDNA viruses suggest that HPV could encode viral miRNAs, which have not been validated thus far. The aim of this thesis was to study the functions of the host miRNAs in HPV 16 oncogene induction and identify novel HPV encoded viral miRNAs. We utilized microarray technology to investigate the effect of E5 on host miRNAs and mRNAs expression in 0–96 hours after E5 induction in a cell line model. Among the differentially expressed cellular miRNAs, we further validated the expression of hsa-mir-146a, hsa-mir-203, and hsa-mir-324-5p and some of their target genes in a time series of 96 hours of E5 induction. Our results indicate that HPV E5 expression has an impact through complex regulatory patterns of gene expression in the host cells, and part of those genes is regulated by the E5 protein. Second, high throughput sequencing was used to identify virus-encoded miRNAs. We prepared small RNA sequencing libraries from ten HPV-associated cervical lesions, including cancer and two HPV-harboring cell lines. For more flexible analysis of the sequencing data we developed miRSeqNovel, an R based workflow for miRNA sequencing data analysis, and applied it to the sequencing data to predict putative viral miRNAs and discovered nine putative papillomavirus encoded miRNAs. Viral miRNA validation was performed for five candidates, four of which were successfully validated by qPCR from cervical tissue samples and cell lines: two were encoded by HPV 16, one by HPV 38, and one by HPV 68. The expression of two HPV 16 miRNAs was further supported by *in situ* hybridization, and colocalization with p16INK4A staining, a marker of cervical neoplasia. Prediction of cellular target genes of HPV 16 encoded miRNAs suggests that they may play a role in cell cycle, immune functions, cell adhesion and migration, development and cancer, which were also among the functions targeted by the E5 regulated host cell mRNA and miRNAs. Two putative viral target sites for the two validated HPV 16 miRNAs were mapped to the E5 gene, one in the E1 gene, two in the L1 gene, and one in the long control region (LCR).

ABBREVIATIONS

5'	five prime
3'	three prime
cDNA	complementary DNA
CGH	comparative genomic hybridization
ChIP-on-chip	on-chip chromatin immunoprecipitation
ChIP-Seq	chromatin immunoprecipitation followed by next generation sequencing
CIN	cervical intraepithelial neoplasia
CLIP-Seq	crosslinking immunoprecipitation-high-throughput sequencing
Cy3	cyanine 3
Cy5	cyanine 5
DAVID	Database for Annotation, Visualization and Integrated Discovery
DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
FDR	false discovery rate
FFPE	formalin-fixed paraffin-embedded
GEO	Gene Expression Omnibus
GO	Gene Ontology
HaCaT	immortal human keratinocyte cell
HeLa	immortal human cervical cancer cell
HPK II	HPV 16 immortalized human keratinocyte cell
HPK IA	HPV 16 immortalized human keratinocyte cell
HPV	human papillomavirus
iCLIP	individual nucleotide resolution CLIP
KEGG	Kyoto Encyclopedia of Genes and Genomes
LCR	long control region
logFC	log2 fold change
MCF7	immortal human breast cancer cell
miRNA	microRNA
miRNA*	star microRNA
miRNP	microRNA ribonucleoprotein complex
mRNA	messenger RNA
NCBI	National center for biotechnology information

NGS	next generation sequencing
nt	nucleotide
ORF	open reading frame
p53	tumor protein 53
PAR-CLIP	photoreactive nucleotide-enhanced crosslinking and immunoprecipitation
pRb	retinoblastoma protein
pri-miRNAs	primary miRNAs
pre-miRNAs	precursor miRNAs
pol II	RNA polymerase II
RNA	ribonucleic acid
RBP	RNA-binding protein
RNA-Seq	Transcriptome sequencing
SCC	squamous cell carcinoma
SMRT	single molecule real time
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SOLiD	Sequencing by Oligonucleotide Ligation and Detection
SV40	simian virus 40
U2OS	immortal human osteosarcoma cell
UTR	untranslated region

1. Introduction

1.1 Human Papillomavirus (HPV)

Human papillomaviruses (HPVs) belong to the family of *Papillomaviridae* and they preferentially infect keratinocytes of mucous membranes or skin and cause numerous benign and malignant lesions at different anatomical locations (Doorbar *et al.* 2012). Close to 200 HPV types have been characterized, which are classified into low-risk and high-risk types (Bernard *et al.* 2010). Low-risk HPV types, such as HPV 6 and 11, can cause benign warts or condylomata acuminata (Lorincz *et al.* 1992). High-risk HPV types, such as 16 and 18, can cause squamous epithelial lesions that have potential to become malignant (zur Hausen 2009). HPV infection is considered the most common sexually transmitted disease worldwide. It is not rare to have co-infection with multiple HPV types or sequential infection with phylogenetically related types (Chaturvedi *et al.* 2011). Two commercial prophylactic vaccines have shown success in preventing HPV infections. However, it is noteworthy that they do not work against all cancer-related types, and do not cure acquired infections. There are some challenging aspects limiting HPV research: first, it is difficult to isolate viral particles from tissue samples and, second, HPVs cannot replicate in conventional cell culture systems (Chow *et al.* 2010).

1.1.1 HPV genome

HPVs are a group of small non-enveloped DNA tumor viruses with a double stranded circular DNA genome approximately 8,000 base pairs in size (Zheng *et al.* 2006). HPV genome can be divided, in general, into three major regions: an early region with ORFs, a late region and a long control region (LCR). The early region encodes six nonstructural regulatory proteins (E1, E2, E4, E5, E6, and E7). The late region locates downstream of the early region and encodes L1 (major) and L2 (minor) structural capsid proteins. The LCR region has no protein-coding function, but provides the promoter and multiple transcription factor binding sites involved in the regulation of viral transcription and replication (Zheng *et al.* 2006).

E1 and E2 are involved in viral DNA replication and the regulation of early transcription (Doorbar *et al.* 2012). E4 is considered to function in virion release (Doorbar *et al.* 1997). E5, E6, and E7 are oncogenes and their expression induce cell immortalization and transformation (Doorbar 2006). A characteristic in cervical carcinogenesis by high-risk HPVs is that viral genomes are commonly integrated into the host cell genome, resulting in the loss of functional viral E2 gene (Romanczuk *et al.* 1992).

1.1.2 HPV oncogenes

The most prevalent HPV type found in cervical cancer is HPV 16, which

encodes three oncogenes: E5, E6, and E7, which are associated with the proliferation of suprabasal epithelial cells (Moody *et al.* 2010). The E6 and E7 can bind to and stimulate the degradation of the tumor suppressors p53 (Werness *et al.* 1990) and pRb (Dyson *et al.* 1989). Their oncogenic potentials are largely correlated with these interactions (Heck *et al.* 1992; Nguyen *et al.* 2002), but their interference with the functions of other intracellular proteins plays an important role as well (Balsitis *et al.* 2006; Shai *et al.* 2007). The expression of cyclin-dependent kinase inhibitor p16INK4A (also known as p16 or CDKN2A) is upregulated due to the interaction of E7 and pRb, and can be used as a marker of persistent high-risk HPV infection (Klaes *et al.* 2001). The expression of E6 and E7 proteins can be repressed by a high expression level of E2 protein, so the loss of E2 gene upon integration of the viral genome into the host genome will lead to increased expression of E6 and E7 (Romanczuk *et al.* 1992).

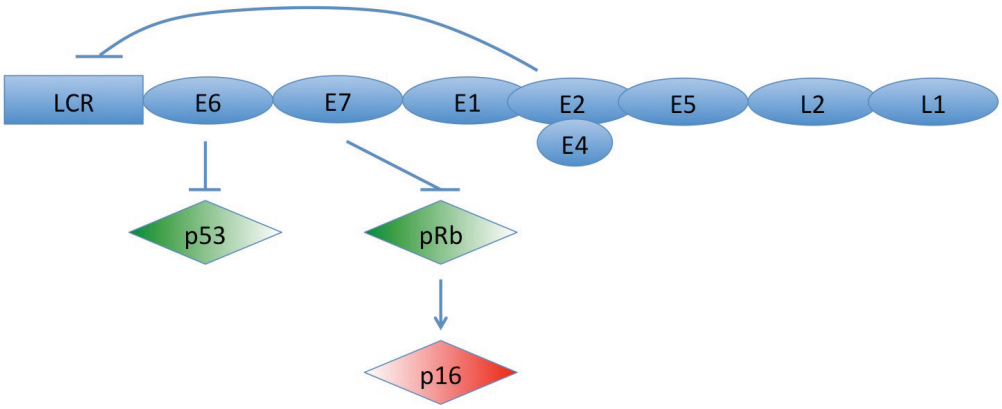

Figure 1. Schematic representation of HPV 16 oncogene activities

HPV 16 E6 and E7 bind and degrade p53 (Werness *et al.* 1990) and pRb (Dyson *et al.* 1989). E7 and pRb interaction also leads to higher expression of p16. E2 can prevent high expression of E6 and E7 (Romanczuk *et al.* 1992). E2 has binding sites in LCR region to regulate E6 and E7 expression.

The role of the E5 oncoprotein in the transformation process is less well studied and understood. The HPV 16 E5 coding sequence is frequently, but not always, deleted when the viral genome is integrated into the host cell genome (Schmitz *et al.* 2012). E5 has been reported to transform tissue-cultured murine fibroblasts and keratinocytes alone (Pim *et al.* 1992; Straight *et al.* 1993) as well as to enhance the immortalization potential of E6 and E7 proteins (Stöppler *et al.* 1996). HPV 16 E5 contributes to tumorigenicity in nude mice (Leechanachai *et al.* 1992) and contributes to skin

carcinogenesis in E5 transgenic mice (Genther Williams *et al.* 2005; Maufort *et al.* 2007). It has been suggested that HPV 16 E5 acts as an oncogene primarily by enhancing the activation of the epidermal growth factor receptor in a ligand-dependent manner (Straight *et al.* 1993; Crusius *et al.* 1997; Crusius *et al.* 1998; Tomakidi *et al.* 2000), but the mechanisms of E5 action have not yet been established, due to a limited number of studies. Very recently it was suggested that E5 alone might have a high oncogenic potential, because E5 transgenic mice were shown to develop cervical cancer after prolonged estrogen treatment (Maufort *et al.* 2010). Additionally, HPV 16 E5 potentiated the effect of E6 and E7 oncogenes in inducing cervical disease (Stöppler *et al.* 1996).

Our group has previously shown that E5 alters the expression of a number of host protein-coding genes in cultured human keratinocytes (Kivi *et al.* 2008). Specifically, we observed that genes implicated in cell motility and cell adhesion are affected by E5 expression. We also showed enhanced motility of E5 expressing cells in an *in vivo* wound healing experiment, which suggests that E5 is implicated in the carcinogenic process (Kivi *et al.* 2008).

Establishment of HPV infection requires the host cell to pass early cell cycle progression and enter M phase (mitosis) in the undifferentiated proliferating basal cell layer (Pyeon *et al.* 2009). Progeny virus production occurs exclusively in differentiated suprabasal layers of the epithelium, and cannot take place if epithelial differentiation is disturbed. Regulation of papillomavirus replication and successful progeny virus production, or pathogenesis of HPV-associated diseases is not completely understood. The environment of epithelial cells committed to differentiate is crucial and essentially involves regulatory changes in mRNA and microRNA expression (Calin *et al.* 2006).

1.1.3 Biology of cervical cancer

Cervical cancer is the third most common cancer in women worldwide (Jemal *et al.* 2011). Cervical HPV infection is a common sexually transmitted infection. It is an evolving process from conversion of normal cervical epithelium to cervical intraepithelial neoplasia (CIN) that is subsequently transformed to invasive cervical cancer (Bodily *et al.* 2011). Squamous cell carcinoma (SCC) of the cervix develops through CIN grades 1-3, which are pre-malignant phases that are well defined by cytological examination: CIN1 (mild dysplasia) to CIN2 (moderate dysplasia) to CIN3 (severe dysplasia/carcinoma *in situ*) (Moody *et al.* 2010). Although most women will at some time be infected with HPV, very few will progress to invasive disease. A proportion of all CIN grades may regress, but CIN3 is considered a pre-cancer with potential to progress to cervical cancer (Schiffman *et al.* 2007). Further, cervical cancer is associated with HPV. Persistent infection by the high-risk HPV types is a prerequisite for cervical cancers, but not sufficient cause of cervical cancer (Woodman *et al.* 2007). The concurrent or sequential detection of different HPV types is common. Two HPV vaccines are currently on the market, both of which protect

against the HPV 16 and 18. Although most HPV infections do not progress to cancers, the high prevalence still makes HPV-related cancers among the most common malignancies.

1.1.4 HPVs and cervical cancer

High-risk HPV infection plays a critical role in the pathogenesis of cervical cancer in women worldwide and is an important etiologic agent in other anogenital cancers (zur Hausen 2002; Woodman *et al.* 2007). High-risk HPV genotypes, including HPV 16, 18, 31 and others, cause almost all cervical cancers and their immediate precursors worldwide (Schiffman *et al.* 2007). High-risk human papillomavirus types 16 and 18 are known to be associated with more than 70% of cervical cancers (zur Hausen 2009; Guan *et al.* 2012).

Although HPVs remain in episomal form in benign, precancerous lesions, they frequently integrate into the cellular genome in cervical cancers, and this integration may play a role in cancer development (Pett *et al.* 2007). Variable portions of the viral genome are present in the integrated genomes, but common features include sustained expression of E6/E7 under the control of their viral promoter as well as loss of the E2 gene (zur Hausen 2000). As described above, HPV E2 is also a transcription repressor that can directly bind to and repress the E6/E7 promoter of high-risk HPVs (Romanczuk *et al.* 1990). The loss of E2 expression causes the oncogene E6 and E7 expression to reduce p53 and pRb expression, which in turn stimulates cellular proliferation. As such, disruption of the E2 gene has been mechanistically linked to malignant progression of HPV-associated cancers. High-risk HPV E6 and E7 proteins are universally expressed in human HPV-positive cancers *in vivo* and function cooperatively *in vitr*o, thus establishing these proteins as the primary viral oncogenes (zur Hausen 2002). The viral E5 protein can further contribute to HPV malignancy (DiMaio *et al.* 2001), and carries an independent carcinogenic potential in transgenic mouse models (Maufort *et al.* 2007; Maufort *et al.* 2010). Deregulated expression of the high-risk HPV oncogenes is a critical event in the progression of HPV-positive lesions (Gao *et al.* 2010).

1.2 microRNA biogenesis and functions

MiRNAs (miRNAs) are short (18-25 nucleotides) noncoding RNAs that post-transcriptionally regulate messenger RNA (mRNA) expression. Lee *et al.* discovered the first miRNA, lin-4 in *C. elegans*, in 1993 (Lee *et al.* 1993), and to date, over 21 thousands miRNAs, including 1,600 human miRNAs, have been identified (miRBase release 19, 2012 (Kozomara *et al.* 2011)). They play important roles in cellular processes such as immune response, cell differentiation and proliferation, metabolism, apoptosis, and oncogenesis (Krol *et al.* 2010). MiRNAs are important players in oncogenesis (Esquela-Kerscher *et al.* 2006). Aberrantly expressed miRNAs can act as oncogenes or tumor suppressor genes.

Although our ability of predicting miRNA targets is still imperfect, it is known

that in animals, the 5' end of an miRNA plays the crucial role in selecting the targets (Lewis *et al.* 2003). The 2^{nd}-8^{th} nucleotides (minimally 2^{nd}-7^{th}) from the 5' end of the miRNA is the so-called "seed" region. The seed region binds to the 3' UTR region of the target mRNAs, leading to degradation of the target mRNA or blocking its translation (Valencia-Sanchez *et al.* 2006). Though most miRNA target sites in animals are found in 3' UTR, miRNAs can also bind either in 5' UTR or coding region (Kumar *et al.* 2012). Many computational algorithms for prediction of miRNA targets are based on this rule (Sethupathy *et al.* 2006). Furthermore, if the potential binding sites in the 3' UTRs are evolutionally conserved cross species, the target mRNAs are considered as conserved targets. Because of short sequences in the seed region, the prediction suggests that one miRNA can regulate even hundreds of target mRNAs and many genes are putative targets of more than one miRNA, which has also been confirmed by experimental validations (Hsu *et al.* 2011). MicroRNA expression is altered in a number of human diseases spanning from psychiatric disorders (Maes *et al.* 2009) to several cancers (He *et al.* 2005). Moreover, they play a major role in regulating host gene expression in various viral infections (Roberts *et al.* 2010).

1.2.1 Human microRNA biogenesis

MiRNAs are found in the human genome as independent loci or within intronic regions of other genes (Cai *et al.* 2004; Kim *et al.* 2007) usually transcribed by RNA polymerase II (pol II) as primary miRNAs (pri-miRNA) (Lee *et al.* 2004). Pri-miRNAs are several kilobases long and may contain one to several hairpin structures for multiple individual mature miRNAs. Pri-miRNAs are cleaved by RNase III-type enzyme in the nucleus, named Drosha, to at least one or several precursor miRNAs (pre-miRNAs). These pre-miRNAs are exported from the nucleus in a process involving the Exportin-5 protein (Kim *et al.* 2009). Intronic pre-miRNAs are generated as a product of splicing of the host gene (Berezikov *et al.* 2007). Each pre-miRNA forms a hairpin structure, which encodes two products, one mature, and one star miRNA (miRNA*). In the cytoplasm, the pre-miRNA hairpins are cleaved by the RNase III enzyme Dicer (Murchison *et al.* 2004), and the mature miRNAs are incorporated into the RNA-induced silencing complex (RISC), where they bind to their targets (Schwarz *et al.* 2003). Typically, miRNA* is less likely to associate with RISC and is usually thought to be degraded based on the lower expression levels (Kozomara *et al.* 2011). Mature miRNAs almost always have variants, named isomiRs, which can also be functional miRNAs (Cloonan *et al.* 2011). The star miRNAs were thought to be non-functional. However, more and more evidence shows that miRNA* can be functional and miRNA* may be highly expressed (Yang *et al.* 2011). As a result, the miRNAs nomenclature is changing from miR/miR* to -5p/-3p, which identifies the miRNA by its location in the 5'/3' arm of the hairpin.

1.2.2 Human microRNA functions

The main molecular function of miRNA is guiding the RISC complex binding target mRNAs, inhibiting translation, and/or cleavage of target mRNA (Volinia *et al.* 2010). Each miRNA can potentially target several mRNAs. Studies have shown that a single miRNA can directly down-regulate the production of even hundreds of proteins either by repression of translation and mRNA degradation (Hsu *et al.* 2011). The seed sequence in the 3'UTR of target mRNA has been confirmed to act as the main motif of miRNA-mediated regulation of protein production (Volinia et al. 2010).

Deregulation of miRNA expression emerges as the main mechanism that triggers the loss or gain of function in cancer cells (Calin *et al.* 2006). With this knowledge, it now seems obvious that aberrant expression of miRNAs is involved in cancer as well as other human diseases (Esquela-Kerscher *et al.* 2006). Multiple miRNAs are deregulated in human disease, supporting the hypothesis that miRNAs are involved in the initiation and progression of disease (Jiang *et al.* 2009). Recent data have shown that miRNA deregulation or mutations correlate with various human diseases, and that miRNAs can function as tumor suppressors and oncogenes (Visone *et al.* 2009).

1.2.3 miRNAs in carcinogenesis and in human cancers

Studies have shown that the overexpression of miRNAs can down-regulate a tumor suppressor or other genes involved in cell growth or differentiation, thereby contributing to tumor formation by stimulating proliferation, angiogenesis, and/or invasion (Zhang *et al.* 2007). miRNAs can act as oncogenes, so-called “oncomirs” (Visone *et al.* 2007; Veronese *et al.* 2010). Similarly, the function of miRNAs in diseases is dependent on their targets for mRNA, where they might act as oncogenes or tumor suppressors. It is interesting to note that some miRNAs may have dual functions as both tumor suppressors and oncogenes (Gebeshuber *et al.* 2009), depending on the context. Specific upregulated or downregulated miRNAs have been shown to correlate with particular tumor types (Volinia *et al.* 2006). Altered miRNA expression profiles have been reported in breast cancer (Iorio *et al.* 2005), cervical cancer (Xie *et al.* 2012), colon cancer (Borralho *et al.* 2011), lung cancer (Hayashita *et al.* 2005), ovarian cancer (Iorio *et al.* 2007), prostate cancer (Nadiminty *et al.* 2012), and thyroid papillary carcinomas (Pallante *et al.* 2006). Moreover, some studies have suggested that miRNAs have the potential as biomarkers for diagnosis, prognosis, and cancer therapeutics in a variety of human diseases (Visone *et al.* 2011).

However, the detailed molecular basis of miRNA-mediated gene regulation is not fully understood and their role in tumor genesis remains largely unknown. The consistent deregulation of several miRNAs in disease will also affect expression of their gene targets. The identification and validation of targets of the deregulated miRNAs is essential to understand the molecular and

biological functions that are influenced as the function of each miRNA is dictated by its own targets (Volinia *et al.* 2010). Because dysregulation of miRNA has been associated with various diseases, the identification of specific regulators of miRNAs will be helpful in developing new therapeutic agents (Shenouda *et al.* 2009).

1.2.4 Cervical cancer and cellular miRNA

Although high-risk HPVs are associated with cervical cancer, a HPV infection alone is not sufficient to induce the malignant transformation of HPV-infected cervical epithelial cells (zur Hausen 2002). The expression of many cellular genes and miRNAs take place in carcinogenesis. Several host miRNA expression profiles in cervical cancer have been explored to identify specific miRNAs that are important for carcinogenesis. Differentially expressed cellular miRNAs in cervical cancer tissue or cervical cancer-derived cells have been previously reported (Lui *et al.* 2007; Wang *et al.* 2008; Ma *et al.* 2012; Gocze *et al.* 2013; Wilting *et al.* 2013). Several upregulated miRNAs have been recognized; including hsa-mir-21 (Yao *et al.* 2009), hsa-mir-127 (Lee *et al.* 2008), hsa-mir-146a (Wang *et al.* 2008; Visone *et al.* 2009), hsa-mir-182 (Tang *et al.* 2013), hsa-mir-199a (Lee *et al.* 2008) and hsa-mir-224 (Shen *et al.* 2013), and they may function as oncogenes. Downregulated miRNAs such as hsa-mir-34a (Wang *et al.* 2009; Li *et al.* 2010; Pang *et al.* 2010; Gocze *et al.* 2013), hsa-mir-124 (Wilting *et al.* 2010), hsa-mir-143 (Lui *et al.* 2007; Wang *et al.* 2008; Lajer *et al.* 2012), hsa-mir-145 (Wang *et al.* 2008; Lajer *et al.* 2012), hsa-mir-200a (Hu *et al.* 2010), hsa-mir-214 (Yang *et al.* 2009) and hsa-mir-218 (Martinez *et al.* 2008; Zhou *et al.* 2010; Yamamoto *et al.* 2013) have been identified and their function as tumor suppressors in cervical cancer carcinogenesis has been suggested.

Downregulation of human miR-218 was specifically addressed to the HPV 16 E6 oncogene in HPV-positive cell lines, cervical lesions, and cancer tissues (Martinez *et al.* 2008; Yu *et al.* 2012). Human miR-218 targets the expression of its target gene LAMB3, which has a role in enhancing cellular migration and tumorigenicity (Zhou *et al.* 2010; Yamamoto *et al.* 2013). Human miR-21 has been identified as a cancer-associated miRNA and is overexpressed in many cancers, including cervical cancer (Lui *et al.* 2007; Wang *et al.* 2008). It was recently shown that inhibition of hsa-mir-21 in HPV 18-containing HeLa cervical cancer cells causes a strong suppression of cell proliferation (Yao *et al.* 2009). Downregulation of hsa-miR143 in cervical cancer cell lines has been reported (Lui *et al.* 2007). It thus seems obvious that, similar to other cancers, miRNAs play an important role in the development of cervical cancer.

1.2.5 Viral microRNA biogenesis

A number of DNA viruses and some RNA viruses encode their own miRNAs as well. Most of the known viral miRNAs are found in herpesviruses (Grundhoff *et al.* 2011). Polyomaviruses (Sullivan *et al.* 2005), adenoviruses, (Xu *et al.* 2007)

and ascoviruses (Hussain *et al.* 2008) are also known to encode their own miRNAs. The precursor miRNAs can be generated by Drosha/Dicer cleavage of primary miRNAs (pri-miRNAs), mirtron, tRNase Z cleavage of tRNA-like pri-miRNA, or alternative folding of transfer RNAs (Reese *et al.* 2010) or small nucleolar RNAs (Skalsky *et al.* 2010). Viral miRNAs are mostly generated from pre-miRNA by Dicer cleavage and incorporated into the RISC, similar to host miRNAs (Grundhoff *et al.* 2011).

To date, no study has been able to identify and validate viral miRNAs in papillomavirus infected cells using either standard sequencing (Cai *et al.* 2006) or next generation sequencing techniques (Lui *et al.* 2007). However, Gu *et al.* (Gu *et al.* 2011) recently reported prediction of several miRNA in mucosal and cutaneous HPVs; but despite considerable efforts, no validated papillomavirus miRNAs have been established so far. The lack of an efficient cell culture system to study viral replication in the context of epithelial differentiation and maturation has complicated miRNA discovery in HPV.

1.2.6 Viral microRNA functions

Viral miRNAs can target both viral and cellular mRNAs for downregulation (Murphy *et al.* 2008). The identification of viral targets of viral miRNAs is often easier due to the smaller genome size and in some cases the viral miRNAs are expressed from transcripts that overlap with the antisense strand of viral genes that they regulate. Potential cellular targets for viral miRNAs could be cellular genes involved in antiviral defense systems, cell proliferation and survival, immune recognition, and stress responses. The viral miRNAs contribute to cellular reprogramming by regulating the switch from latent to lytic viral infection, and by modulating the immune responses in the infected host (Skalsky *et al.* 2010). Polyomavirus miRNAs target viral early transcripts to negatively regulate T antigen expression, and they also promote immune evasion by targeting cellular genes involved in host immune response, which subsequently leads to enhanced viral replication (Seo *et al.* 2008; Seo *et al.* 2009; Bauman *et al.* 2011; Lee *et al.* 2011). All these functions are very important for viruses in the establishment of persistent infections. The simian virus 40 (SV40) encoded miR-S1 is transcribed from the DNA strand harboring the late genes, and is complementary to early T antigen transcripts (Sullivan et al. 2005). The cleavage of T antigen function has been further confirmed; however, it is unessential for replication of virus. On the other hand, downregulation of T antigen transcripts helps virus replication by lowering cytotoxic T cell recognition. Thus viral miRNA probably contributes to the proper balance of viral gene products at specific phases of the infection, most likely during latency. Many viral miRNAs are likely to help viruses cause long-term infection by evading the immune response, which is a requisite for cancer development (Pfeffer *et al.* 2006). Instead of being oncogenic, viral miRNAs could help mediate the survival of viruses from a cellular immune system and thus contribute to tumor genesis. Functional and molecular

similarities among double stranded DNA viruses causing long-term latent infections, especially SV40 (Sullivan *et al.* 2005), human polyomaviruses BKV and JCV (Seo *et al.* 2008), and BPCV (Chen *et al.* 2011), suggest that HPV could also encode viral miRNAs.

1.3 Gene expression profiling technology

High-throughput technologies, especially microarray and next generation sequencing (NGS), are ideal tools to measure gene expression globally and in parallel cross sample to understand the complex functional mechanisms (Wang *et al.* 2009). They have changed the way of genetic research from a few interesting genes to large-scale investigations, usually the whole genome and even metagenomics. Ever since the microarray technology emerged, it has been the favorite platform to achieve this aim (Schena *et al.* 1995).

Although microarrays and NGS are widely used and have many advantages as compared to many traditional technologies, some issues remain to be considered. In the early days of these methods, the quality and reproducibility of results are the primary concern. With the improvement in technology and development of laboratory protocols, the quality of generated data has improved and many meta-analyses have been performed using data from different laboratories (Krol *et al.* 2010).

There is also demand to improve the way the results are interpreted to bring them into meaningful biological context. For instance, the aim of gene ontology (GO) project is to describe every gene in a set of comprehensive controlled terms (Ashburner *et al.* 2000). Besides that, many other databases contribute to the gene annotation, such as KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway (Kanehisa *et al.* 2012). Annotations of model organism genomes are much better than other organisms. However, annotations still contain numerous errors and need further corrections. The DAVID (Database for Annotation, Visualization and Integrated Discovery) is a widely used web-accessible tool to analyze the functional annotations behind lists of large number of genes (Huang da *et al.* 2009).

1.3.1 Microarray

From its first application for transcriptome analysis (Schena *et al.* 1995), microarray technology has been adapted in many research areas such as genotyping, miRNA profiling, SNPs, splicing variant analysis, copy number variation, comparative genomic hybridization, and ChIP-on-chip (Fan *et al.* 2006).

The gene expression microarray can measure the expression of tens of thousands of mRNAs in a single assay, depending on the microarray design. In Agilent and Illumina arrays, the total RNAs are extracted from the biological samples, converted into cDNA, and labeled with fluorescent dyes for hybridization. After hybridization with the probes fixed onto the microarray surface, the fluorescent signal intensities are detected by a laser scanner. The

image file is further analyzed to estimate the probe signal intensities. This represents the raw microarray data, which is merged with the annotation file to build the expression matrix and used for further analysis.

There are mainly two types of microarrays available for expression analysis: spotted microarrays, which deposit pre-synthesized oligonucleotides, or cDNA probes (usually PCR products) on the array surface; oligonucleotide microarray, which directly synthesize probes or “ink-jet printing” on the surface. The most popular commercial microarray platforms are Affymetrix (multiple shorter probes target each gene), and Agilent (longer and more specific probes (Hughes *et al.* 2001)). In Agilent microarray, samples can be labeled with up to three different dyes and hybridized to the same area in the microarray (Pessa *et al.* 2010).

There are many options for the bioinformatics analysis of microarray data. In general, the following steps are needed: 1) Preprocessing and quality control. The intensities in the scanned pictures need to be converted into numbers for following steps. The overall data reliability can be improved by checking the quality of the microarray image and flagging out low quality spots. 2) Normalization of the data. Normalization is necessary to remove the unavoidable experimental variations, such as differences in labeling efficiencies. 3) Identification of differentially expressed genes. The data are fitted into statistical models to test the differential expression levels. The outputs of these analytical steps are usually tables of genes with related p-values and log2 fold changes (logFC). FDR methods, such as Benjamini–Hochberg method (Benjamini *et al.* 1995), can lower the false positive rate of the results. 4) Functional annotation analysis. This is done by applying functional annotation analysis tools, such as DAVID, to the differentially expressed gene lists. And the result will be a list of functional annotations, which explain the cellular functions of differentially expressed genes and give clues to their biological interpretation (Qian *et al.* 2012).

Compared to NGS technologies, microarrays have some disadvantages in finding novel transcripts and novel variation because of limitations in probe design (Wang *et al.* 2009). Moreover, the raw data in microarray analysis are inferred from hybridization intensity, unlike NGS where it is directly measured. On the other hand, bioinformatics analysis methods for microarrays are more mature than for NGS, which is still in its developing stages. Another advantage of microarrays over NGS is that for low expressed genes, a very large number of reads in NGS is needed to perform reliable statistical analysis (Dillies *et al.* 2012).

1.3.1.1 mRNA and miRNA microarrays

Gene expression profiling is probably the most widely used microarray application (Duggan *et al.* 1999). It is used to explore the expression of a large number of genes, often the transcriptome, in order to identify differentially

expressed genes in different conditions, such as different cells or treatments, or in diseases, such as cancers. However, it also has limitations to investigate other essential biological questions, such as posttranscriptional regulation and modification of mRNAs.

The miRNA expression profiling is similar to mRNA but it is affected by the unique features of miRNAs, such as the short length, low abundance and unequal melting temperatures of mature miRNA (Thomson *et al.* 2004). It is important to distinguish between mature miRNAs and their pri-miRNA and pre-miRNAs. Because of the higher sequence similarity of some miRNAs, especially among some miRNA families, it is crucial to distinguish between molecules where the sequence differs by a single nucleotide. The miRNA microarray can be used for quantification of hundreds or thousands miRNAs in cell lines or tissues samples. A miRNA expression profiling study could be a useful tool to evaluate a detailed molecular analysis of its expression in the process of cancer development.

1.3.2 Next Generation Sequencing

Compared to the traditional Sanger technology, NGS gives much greater speed and throughput at much lower costs per base. The NGS approach altogether omits the cloning step that creates bias and is very labor intensive and expensive when a larger number of reads is needed, for example, *de novo* genome sequencing, transcriptomics, and metagenomics. The most popular NGS platforms are Roche 454, Illumina (former Solexa), and Applied Biosystems (AB) SOLiD. In the 454 system, fragments of DNA are ligated with adapters and then bound to beads. Then DNA are amplified in the emulsion PCR on bead surface. Each bead with amplified fragment is then positioned in a picotiter plate, which further undergo pyrosequencing. The emitted light from synthesis during sequencing is detected as nucleotide signal (Wheeler *et al.* 2008). The Illumina platform uses a solid-phase bridge PCR technology to amplify DNA fragments hybridized via adapters on a slide. Thereafter sequencing these amplified clusters by DNA polymerase using cleavable nucleotide analogs with different fluorescent labels. The emitted light and its position are recorded (Bentley *et al.* 2008). The ABI SOLiD also uses emulsion PCR on beads for amplification of the target libraries, but utilizes ligations for the actual sequencing. Specifically, sequential annealing and ligation of fluorescent labeled oligonucleotide probes are lighted to the DNA fragment immobilized on a glass support, The based are no longer sequentially interrogated, but at spaced intervals. The complete sequence is generated from overlapping of series of ligations (Valouev *et al.* 2008). The output of every sequencing technology is generally called read, which contains the information of the sequence identifier, sequence letters (bases), or numbers and quality values. The read length of 454 is longer, from 100bp (first model) to 1,000bp (latest model), while the read length of Illumina (HiSeq) and SOLiD is about 50-150bp. Moreover, the read format from SOLiD is different from the

other two, usually called color space, while base space is used for 454 and Illumina. Depending on different sample library preparation methods, there are single-read, pair-end, or mate-pair sequencing, which allow different length of insert sizes.

NGS technology provides opportunities for deeper studies of genetics and potential clinical application (Ozsolak *et al.* 2011). In cancer genome, common mutations have numerous forms, including and not limited to SNVs, small insertions and deletions, copy number changes, alternative splicing, DNA rearrangement, and chromosome structural variants (Hawkins *et al.* 2010). RNA-Seq can be used in NGS to sequence cDNA of transcriptome and reveal various aspects of dynamic transcriptome (Nagalakshmi *et al.* 2008).

The epigenetics, including DNA methylation, histone modifications, and chromatin remodeling can also be studied by NGS through bisulfite sequencing and ChIP-Seq. Bisulfite sequencing discovers methylated nucleotides by bisulfite treatment of DNA (Krueger *et al.* 2012). ChIP-Seq combines chromatin immunoprecipitation with NGS to identify the binding sites of DNA-associated proteins. ChIP-Seq is widely used for finding transcription-factor binding sites, histone modifications, and also some of the DNA methylation methods (Gerstein *et al.* 2012).

Besides genome and epigenetics, posttranscriptional RNA editing, including base insertion, deletion, and substitution is common in transcriptome (Li *et al.* 2009). RNA editing will change the exon sequence and may lead to alteration in amino acids. The miRNA also post-transcriptionally regulates gene expression. It is likely that cellular processes are tightly regulated at the genetic, epigenetic, and transcriptome levels that interact in a complex network (Mattick 2004). NGS technology is advantageous for integrated studies of the genome, epigenetics, and transcriptome, which reveal biological mechanisms at multiple levels. The progress of NGS also raises challenges in bioinformatics to analyze such large-scale data (Pop *et al.* 2008).

NGS technology is continuously evolving by updating sample preparation, library protocols, and developing design of sequencers. Improvements in sequencing technologies and bioinformatics analysis tools will lower the cost, enhance the sensitivity and accuracy of detecting variances. Furthermore, the third generation sequencing technology is now also commercially available, for example PacBio RS II from Pacific Biosciences (Eid *et al.* 2009). It is based on single molecule real time (SMRT) sequencing technology. It has the unique feature in producing much longer sequencing reads, and detecting DNA modifications (Feng *et al.* 2013).

1.3.2.1 miRNA sequencing

The miRNA sequencing technology (miRNA-seq) has given new opportunities to quantify the expression of known miRNAs, discover nucleotide variations (including RNA editing, 3’ and 5’ modifications) as well as to predict novel

miRNAs. Compared to miRNA microarray, miRNA-seq offers the advantage of finding novel miRNAs. Moreover, it avoids the probe design challenges caused by miRNA sequence characteristics. Furthermore, variants of known miRNA, which may be caused by RNA editing, can change their sequences or expression levels and thus alter their ability to regulate gene expression (Wu *et al.* 2008). The known miRNA annotations from miRBase are required to find differentially expressed miRNAs (Kozomara *et al.* 2011). It is known that isomiRs are common for miRNAs and the functional isomiRs may not be annotated in miRBase (Kuchenbauer *et al.* 2008; Morin *et al.* 2008).

As the deep sequencing generates millions of reads, the real challenge is now about how to adequately analyze and interpret data. The basic steps behind the analysis of miRNA-seq can be summarized in three stages: (i) reads preprocessing (trim adaptors, quality check, etc.) and mapping to reference genomes; (ii) analysis of known miRNAs, such as differential expression test, isomiR discoveries; and (iii) prediction of novel miRNA on the basis of negative free energy and structure. The results of second and third stages are heavily dependent on the quality of the first stage. The mapping of miRNA-seq has its own bioinformatics challenge because the length of miRNA is short. The sequencing output is longer and needs trimming. Also the methods and parameters of mapping effects the results of following steps significantly (Williamson *et al.* 2013). Current analysis tools are designed to be easy to use, but with loss of flexibility to users to choose methods and decide the parameters. This means the result and performance of the analysis tools are heavily dependent on the data (species) and it is very difficult to assess the results from different tools (Li *et al.* 2012). There are still other issues that can be used for the improvement of current analysis tools such as supporting colorspace reads. Based on the above facts, we find the existing methods are not flexible enough and need a more adjustable workflow for processing miRNA-seq data.

2. Aims of the study

The aim of this project was to study human miRNA functions in HPV oncogene E5 induction by microarray and identifying HPV miRNAs by NGS with suitable bioinformatics analysis. The specific objectives are:

- Evaluating the E5 effects on human epithelial cells in the time series (0-96h) in mRNA and miRNA microarray.

- Designing flexible analysis workflow for miRNA sequencing data.

- Discovering novel HPV miRNAs from tissue samples by high throughput sequencing.

3. Methods

Summary of methods used in this thesis:

Microarray expression analysis (I)
- Human miRNA and mRNA microarray preprocessing (I)
- Differentially expressed miRNA/mRNA analysis (I)

Functional annotation analysis of gene lists (I, III)
Prediction of human miRNA targets (I)
miRNA sequencing data analysis workflow design (II)
miRNA sequencing by SOLiD platform (III)
Mapping and preprocessing the SOLiD sequencing data (III)
Prediction of candidate viral miRNAs (III)
Prediction of novel viral miRNA targets (III)

Dataset analyzed in this thesis:

NCBI GEO database	
miRNA/mRNA microarray data (I)	GSE24908
Public miRNA sequencing data (III)	
Human SOLiD dataset	GSE22145
Human Solexa dataset	GSE16579
Arabidopsis thaliana Solexa dataset	GSE19694
HPV miRNA sequencing data (III)	GSE42380

3.1 Microarray analysis (I)

Microarray slides were scanned by Axon GenePix 4200 AL scanner (Molecular Devices, Downington, PA). Images were analyzed by the software GenePixPro® 6.0 (Molecular Devices) and the output data were then imported into R software (R Core Team 2013) and preprocessed by the Bioconductor package Limma (Smyth 2004). For finding the differentially expressed genes, empirical Bayes test was utilized (nominal p-value < 0.001), and for miRNA, FDR model was applied (p-value < 0.01 after Benjamini-Hochberg correction) between E5-expressing and control cells in each time point in Table 1.

Table 1. Summary of miRNA and mRNA experiment design in time series.

	0h	2h	4h	24h	48h	72h	96h
miRNA microarray	x		x	x	x	x	
mRNA microarray	x	x	x	x	x	x	x

3.2 Functional annotation analysis of gene lists (I, III)

Functions of interesting genes were analyzed by the DAVID 6.7 annotation tools (Dennis *et al.* 2003; Huang da *et al.* 2009) to identify related and

interesting biological functions. Default DAVID parameters were utilized. In Paper I, the genes are an intersection of predicted targets and differentially expressed genes. In Paper III, the genes are predicted targets of miRNAs.

3.3 Prediction of cellular microRNA targets (I)

Putative targets of each human miRNA were predicted by using eight widely used algorithms in default settings for human miRNAs:
DIANA – microT (Maragkakis *et al.* 2009)
miRanda (Betel *et al.* 2008)
miRDB (Wang *et al.* 2008)
miRWALK (Dweep *et al.* 2011)
PicTar (Krek *et al.* 2005)
PITA (Kertesz *et al.* 2007)
RNA22 (Miranda *et al.* 2006)
TargetScan (Grimson *et al.* 2007).

3.4 miRNA sequencing data analysis workflow design (II)

The detailed processing schemes of the miRNA sequencing workflow, which we named miRSeqNovel, were described in Paper II (Figure 1 in Paper II). In general, workflow is designed in R and contains three major parts: data preprocessing, known miRNA analysis, and novel miRNA prediction. This workflow integrated some available methods/software into its pipeline and also tried to be extendable to other methods. For data preprocessing, the mapped result in SAM/BAM formatted output (Li *et al.* 2009) from Illumina/Solexa data can be read into R. For known miRNAs annotated in miRBase, the read counts for each miRNA will be extracted to build new tables, which can be directly used to find differentially expressed miRNAs as well as isomiRs expressions. For predicting novel miRNAs, there are two strategies to find the pre-miRNA sequences: to look for pairs of mapped reads within a certain range of each other and for a mapped read whose distance from its closest mapped read is longer than the defined gap, the genomic regions upstream and downstream of the reads are searched for possible sequences forming a hairpin secondary structure. We applied the RNAfold program in ViennaRNA (Hofacker 2003) to validate the hairpin structures of the predicted pre-miRNAs and only keep those with similar character to know miRNAs. All cited software packages are listed here:
RNA2MAP (now part of Lifescope software (Life Technologies 2013))
MaToGff (now part of Lifescope software (Life Technologies 2013))
Limma (R/Bioconductor package (Smyth 2004))
Biostrings (R/Bioconductor package (Pages *et al.* 2013))
GenomicRange (R/Bioconductor package (Aboyoun *et al.* 2013))
edgeR (R/Bioconductor package (Robinson *et al.* 2010))
DESeq (R/Bioconductor package (Anders *et al.* 2010))
RNAfold (part of ViennaRNA (Lorenz *et al.* 2011))

3.5 miRNA sequencing by SOLiD platform (II)

A total of 12 samples: one of HPK IA cells, one of HPK II cells (Dürst *et al.* 1987) and ten FFPE tissue samples, were used for preparing the small RNA libraries and subsequently ligated to adaptors of the SOLiD™ Total RNA-Seq kit (Ambion, Austin, TX). Next, small RNAs were reverse transcribed into cDNA libraries, size selected, and amplified with PCR primers containing the barcodes. Libraries were prepared for emulsion PCR (emPCR) according to SOLiD sequencing instructions followed by sequencing using the SOLiD 4 instrument (Life Technologies, Carlsbad, CA).

3.6 Mapping and preprocessing the SOLiD sequencing data (III)

We used SOLiD small RNA pipeline (RNA2MAP) with default parameters to map the SOLiD colorspace reads that included quality filtering and adapter trimming steps. The viral reference genome was constructed by concatenating 393 complete papillomavirus genomes retrieved from NCBI, including known subtypes and isolates of HPV and papillomaviruses in other species (Supplementary Table S1 in Paper III). Mapped results in Ma format were converted to GFF files using MaToGff (Applied Biosystems), which was further used as an input for miRSeqNovel. The SOLiD raw data is publicly available at GEO, series record GSE42380.

3.7 Prediction of candidate viral miRNAs (III)

We predicted novel virus-encoded miRNA candidates from the small RNA sequencing data using miRSeqNovel's novel miRNA prediction function (Paper II). Due to lack of knowledge of HPV miRNA, we performed the prediction by using human and other miRNA characters. The prediction was done twice in two sets of parameters. In the first round, reads with counts less or equal to three in each library were considered as background and not used for prediction. To predict pre-miRNA candidates in miRSeqNovel, 40nt was used as the gap length for strategy one and 100nt was used as the upstream or downstream length for strategy two. Candidates shown at least in two libraries were further studied for their structures. We selected the candidate miRNAs for further validation based on the first round of predictions. In the additional round, reads from all 12 libraries were pooled together, and then reads with counts less or equal to two were discarded. Subsequent steps were performed similarly as in the first run, except that pre-miRNAs were extended to cover the reads overlapping with the predicted region.

3.8 Prediction of novel viral miRNA targets (III)

Human target genes of novel viral miRNAs were predicted using TargetScan custom miRNA prediction methods (Lewis *et al.* 2005). Putative targets within the viral genome were predicted using TargetScan Perl script.

4. Result

4.1 E5 effects on human miRNA expression

The expression of human cellular miRNA was studied in HPV 16 E5 containing HaCaT cells as compared to controls by Agilent human miRNA microarray. The miRNA expression was considered significantly changed if the p-value was <0.01 after FDR correction (Benjamini-Hochberg) (Table 1 in Paper I). In total, 13 differentially expressed miRNA were shown in 0, 24, 48 and 72 hours after E5 induction in microarray analysis (Table S3 in Paper I). They are hsa-mir-146a (up), hsa-mir-203 (down), hsa-mir-324-5p (down), hsa-mir-200c (up), hsa-mir-214 (up (Peng *et al.* 2012)), hsa-mir-624 (up), hsa-mir-539 (up), hsa-mir-433 (up), hsa-mir-23b (down, previously reported in (Au Yeung *et al.* 2011)), hsa-mir-30a-5p (down), hsa-mir-19a (down), hsa-mir-106a (down), and hsa-mir-107 (down). All of them were further validated by qPCR at the same induction time point and one more at four hour (Table 2 in Paper I and Figure 2). From those, we selected hsa-mir-146a, hsa-mir-203, and hsa-mir-324-5p for further investigation based on their biological relevance. The hsa-mir-146a was constantly upregulated; while the hsa-mir-324-5p was constantly downregulated in E5 inducted cells at all the time points. And the hsa-mir-203 remained unchanged during the initial 24 hours of the experiment but was repressed at later stages (48 and 72 hours).

4.2 Integration of differentially expressed mRNAs and miRNAs

Similar to time series microarray studies of miRNA, the expression of mRNAs was also analyzed in HPV 16 E5 induced HaCaT cells compared to controls by Agilent human mRNA microarray (Table S1 in Paper I). In addition to time points used in miRNA studies, three more time points (2, 4 and 96h) were applied to mRNA study. The mRNA expression was considered as significantly changed by the threshold of modest p-value <0.001 to minimize the type I errors (Table 1 in Paper I). qRT-PCR (32 genes) and western blotting (6 proteins) validation were performed with at the same time point of mRNA microarray with one more at 36 hours (Table 3 in Paper I).

Next, putative microRNA targets were predicted using eight distinct algorithms (see methods) for chosen miRNAs. Then the target lists were intersected with different expressed gene lists, resulting in 101, 79, and 176 putative targets of hsa-mir-146a, hsa-mir-324-5p, and hsa-mir-203, respectively (Figure 1 and Table S4 in Paper I).

Functional annotation analysis results of these targets revealed that they were enriched in cell adhesion, and cell cycle (hsa-mir-146a); cell junction, cell migration, and cell motility (hsa-mir-203); cell death, and cell adhesion (hsa-mir-324-5p) (Table S5 in Paper I).

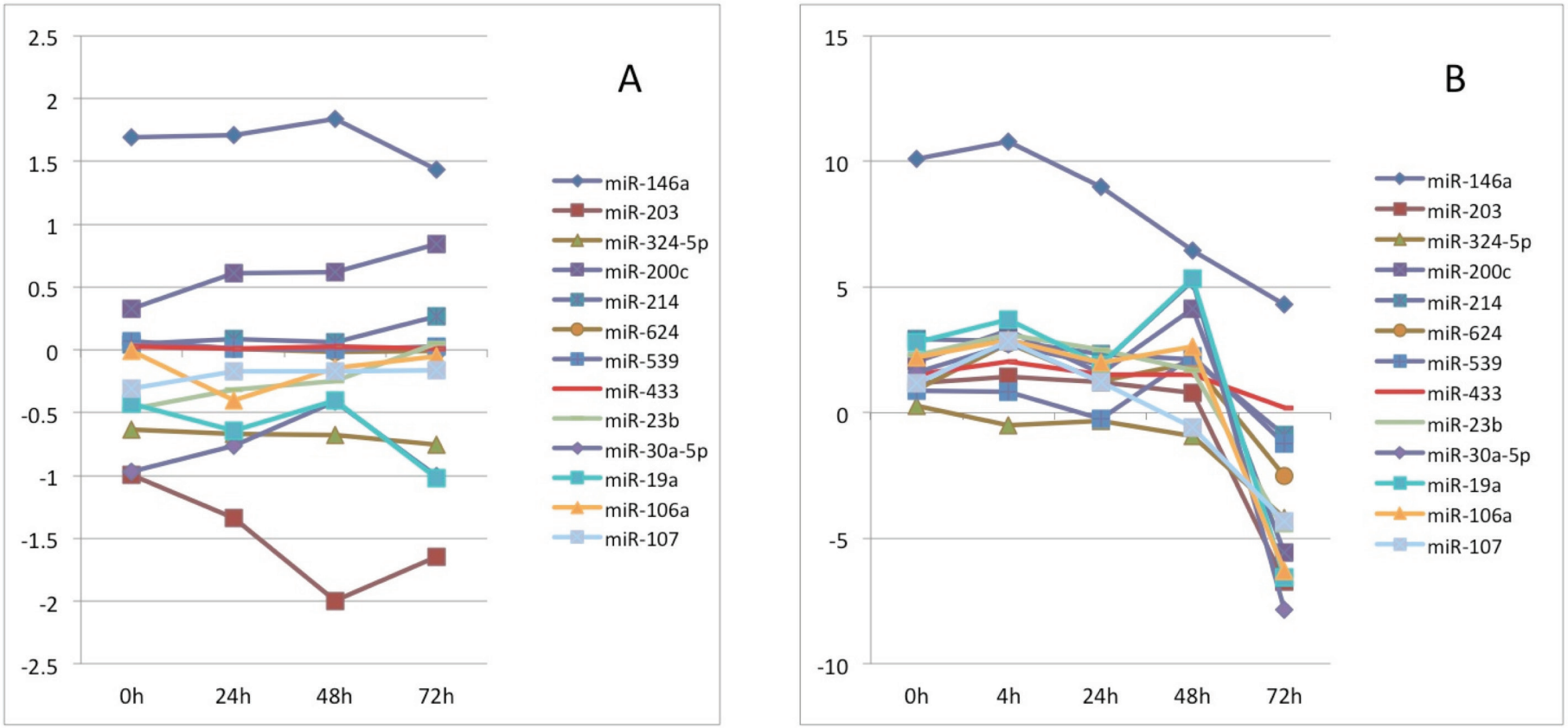

Figure 2. Expression plot of differentially expressed miRNAs in time series

The log2 fold changes of every differentially expressed miRNAs are plotted to the measured time points (0-72h). Note: Some expression plots are not considered as significant different because of high p-values, but still plotted to show trends. A, results of 13 miRNA expressions by microarray. B, result of 13 miRNA expression by qPCR.

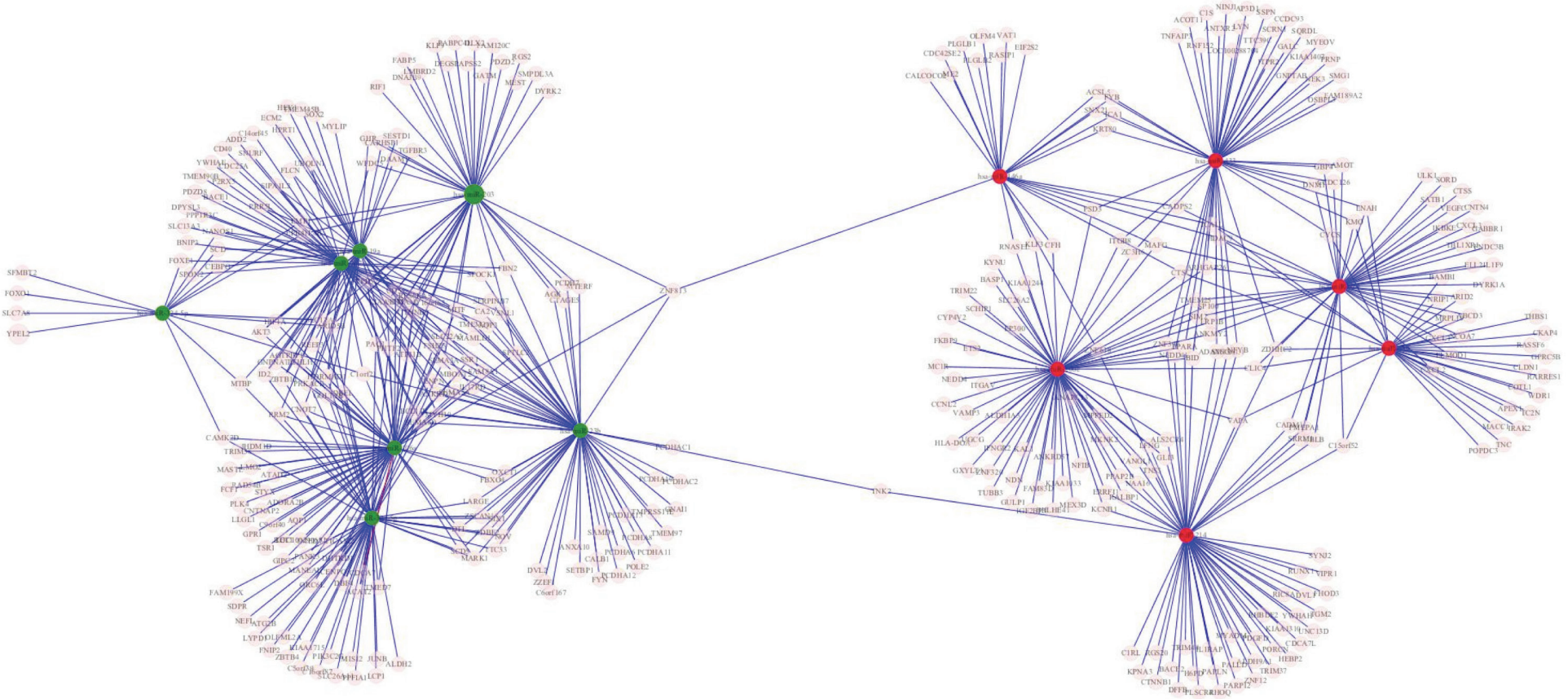

Figure 3. Correlation of miRNA and its negatively correlated targets.

The downregulated miRNAs in E5 expressing cells are marked in green, and upregulated miRNA in red. Negatively correlated and differentially expressed genes are also predicted targets of the connected 13 miRNAs.

To find miRNA involved regulation networks in HPV 16 E5 induction, we compared the expression trend of the differentially expressed miRNA and their predicted targets in differentially expressed mRNAs. We assumed that functional miRNAs and their regulated targets expression would be negatively correlated in the time points. The Figure 3 plots all negatively correlated miRNA-mRNA pairs for all the 13 miRNAs.

4.3 miRSeqNovel workflow and its performance on public data

As the sequencing reads are usually longer (35nt, SOLiD4) than mature miRNA lengths (18-22nt), the reads need trimming first. Then reads can be mapped by any preferred and well-documented mapping software, e.g. RNA2MAP/ Lifescope , Bowtie/ Bowtie2, (Langmead *et al.* 2009; Langmead *et al.* 2012) or BWA (Li *et al.* 2009). This ensures that users can optimize the preprocessing (trim adapters, quality filtering) and mapping without any limitations and utilize sequencing data from multiple platforms, including SOLiD ("csfasta" format), and Illumina/Solexa ("fastq" format) platforms, against any reference genome. Before predicting novel miRNAs, reads mapped to known noncoding RNAs and exon regions are filtered according to Ensembl annotation (Flicek *et al.* 2011) (optional step). The remaining reads will be used as the input for novel miRNA prediction. Next, miRSeqNovel screens mapped reads to find candidate miRNA precursor sequences by checking their secondary structures. By assigning different sets of predictive parameters optimized for animal or plant genomes, we demonstrated that miRSeqNovel could successfully predict most known miRNAs and find conservative novel candidates.

We tested miRSeqNovel on three SOLiD/Solexa miRNA sequencing datasets of human/Arabidopsis samples from NCBI GEO datasets. The results, script used, input files, and user manual are available online (http://sourceforge.net/projects/mirseq/files/). To discover differentially expressed miRNAs, many statistical methods could be applied to the expression, presented by read counts, of known miRNAs annotated in the miRBase database (Kozomara *et al.* 2011). miRSeqNovel by default uses edgeR (Robinson *et al.* 2010) and DESeq (Anders *et al.* 2010), both of which are widely used Bioconductor packages for analyzing sequencing data. Moreover, miRSeqNovel-built count tables could also be used in other methods that might use Bioconductor/R. In human SOLiD dataset (GSE22145, SOLiDv2, 35nt, 6.6M reads) (Marcet *et al.* 2011), we successfully found hundreds of differentially expressed known miRNAs (miRBase V16) by edgeR: 882 out of 2854 annotated mature miRNAs, 172 out of 267 most abundant isomiRs, and 481 out of all 812 common isomiRs. We evaluated the novel miRNA finding function of miRSeqNovel on a set of Solexa miRNA sequencing data from human cancer cell lines HeLa, MCF7 and U2OS (GEO accession number GSE16579, Illumina Genome Analyzer, 35nt, 14.1M reads) (Mayr *et al.*

2009). miRSeqNovel successfully predicted 692 out of 882 (78.5%) known miRNAs expressed in this library. Additionally, miRSeqNovel predicted the presence of 418 novel candidate miRNAs. However, 190 previously reported candidate miRNAs were not predicted by miRSeqNovel. This could either be because the predicted hairpins are not the ones annotated in the miRBase, or because the matching percentage between the two paired sequences is lower than the given threshold. It is important to note that many of these 190 miRNAs were originally predicted by other methods but they have not yet been experimentally validated (Jima *et al.* 2010; Stark *et al.* 2010; Witten *et al.* 2010). For plant data, we used *Arabidopsis thaliana* (GSE19694, Illumina Genome Analyzer, 35nt, 3.1M reads) (Zhang *et al.* 2010) Solexa miRNA sequencing dataset. We successfully found 141 out of 180 (78.3%) known miRNAs presented in this data, along with an additional 371 candidates found by another set of parameters optimized for plant.

4.4 SOLiD sequencing of HPV samples

The total number of reads from SOLiD deep sequencing in HPK IA and HPK II cell lines and FFPE tissue samples varied from 7 million to 97 million, altogether 500 million reads (Table 1 in Paper III). The number of total reads mapped by RNA2MAP to the papillomavirus reference genome ranged from 61 thousands to 1.4 million, roughly 1~2 % in each library.

4.5 Prediction of novel viral miRNA candidates

The mapping result from sequencing was used to predict novel viral miRNA candidates in miRSeqNovel (Paper II). We performed two rounds of predictions to determine whether the pre-miRNA should be extended to cover low count reads (see Methods). Putative HPV miRNAs for further studies were selected based on the clinical relevance of HPV types, particularly HPV 16. The results were finally combined, which revealed a total of nine novel putative HPV-encoded miRNAs (Table 2, Supplementary Table S7, Supplementary Fig.S2, Supplementary Fig.S3 in Paper III). The pre-miRNA sequences were named following the nomenclature instructions in miRBase (Kozomara *et al.* 2011), such as HPV16-mir-H1 for pre-miRNA, and HPV16-mir-H1-1 for mature miRNA.

The HPV type that is most important clinically, HPV 16, encoded five of the selected candidates. Two of the candidates were studied in more detail: HPV16-mir-H1, encoded by a region within the E1 gene on the positive DNA strand, and HPV16-mir-H2, encoded by the negative strand complementary to the LCR (Figure 4). Interestingly, HPV16-mir-H2 coding sequence was found in HPV 16 isolates, but there is a one-nucleotide deletion in the mature miRNA sequence in the prototype HPV 16 genome (NC_001526.2).

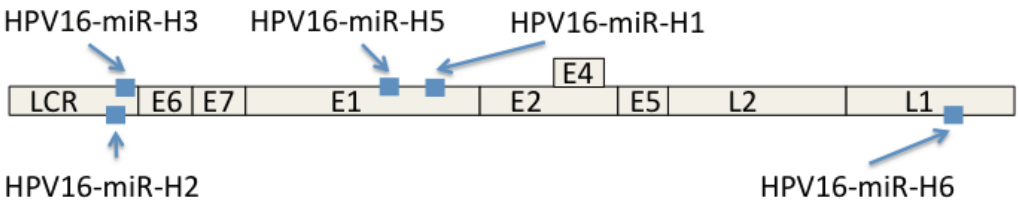

Figure 4. Locations of five predicted HPV 16 novel miRNAs

Locations of HPV16-mir-H1-1 and HPV16-mir-H2-1 in the viral genome are shown as blue bars in Figure 4. The mature HPV16-mir-H1-1 miRNA is located in the loop region of HPV16-mir-H1 pre-miRNA as suggested by miRSeqNovel prediction (Supplementary Table S7 and Supplementary Fig.S4 in Paper III). Other isomiRs in the HPV16-mir-H1 are also found in the loop region. On account of the low read counts in sequencing data, it was challenging to distinguish mature miRNA and star miRNA from RNA degradation background. Because of the background, we further checked whether the predicted miRNA is correct by screening this with VMir (Sullivan *et al.* 2005), and found that the mature HPV16-mir-H1-1 miRNA, not the isomiRs, is located in the arm region of a shorter precursor sequence (Supplementary Fig.S4 in Paper III).

HPV16-mir-H3 is localized in the LCR, HPV16-mir-H5 in the E1 coding region, and HPV16-mir-H6 in the antisense strand complementary to L1 (Figure 4). HPV6-mir-H1 is localized in the antisense strand of HPV 6 E1 region. HPV38-mir-H1 coding sequence is harbored in antisense of E7 region. The coding sequence for HPV45-mir-H1 is located within the L1 region. HPV68-mir-H1 is localized in HPV 68 L1 coding region.

Five HPV miRNA candidates were further validated by TaqMan MicroRNA assays (Method in Paper III) for putative mature miRNAs. Four out of the five miRNAs were presented in cell lines and tissue samples (Table 3, Table S8 in Paper III). All predicted viral microRNAs were detected at high cycle counts in qPCR, suggesting fairly low expression levels in tissue samples. HPV16-mir-H1-1 and HPV16-mir-H2-1 coding regions were further confirmed by PCR amplification of the relevant HPV regions followed by Sanger sequencing for a representative set of samples in cell lines and several tissue samples (Table 3 in Paper III).

HPV genotypes in the tissue samples were identified by a sensitive assay recently developed by our group (Ritari *et al.* 2012), as well as a Luminex based genotyping assay (Method in Paper III). As expected, HPV types 16 and 18 were the most frequently occurring genotypes, and several samples harbored two HPV types (Table 3 in Paper III).

We performed *in situ* hybridization (Method in Paper III) in tissue samples for HPV16-mir-H1-1 and HPV16-mir-H2-1 (Table 4 in Paper III). Cytoplasmic

signals for HPV16-mir-H1-1 were detected in several tissue samples, often colocalizing with p16INK4A (Figure 2 in Paper III), which is a surrogate marker for high-risk HPV. The encoding region for HPV16-mir-H1 is located within the E1 gene, a probe specific for E1 mRNA was used as the control in altogether ten tissue samples (Figure S5 in Paper III). The negative signals of the E1 mRNA probe confirmed the specificity of our hybridization conditions. The HPV16-mir-H1-1 signal in our analysis also localized differently from a previously published work on E1 mRNA signal (Coupe *et al.* 2012). Low level expression of HPV16-mir-H2-1 was detected in one SCC tissue (Figure 2 in Paper III). Human miR-205 was used as a positive control for miRNA expressed in cervical epithelium (Xie *et al.* 2012), which also showed a different pattern between disease and healthy tissues.

4.6 Prediction of viral miRNA targets

We searched for targets of HPV16-mir-H1-1 and HPV16-mir-H2-1 by TargetScan (Grimson *et al.* 2007) . Both of these miRNAs have a unique target fingerprint of 137 and 176 genes, respectively, in the human genome (Supplementary Fig.S8, Supplementary Table S9). Interestingly they share 15 predicted targets: CDC2L6, EIF2C1, IMPAD1, BNC2, SNX27, TNRC6B, BACH2, CYP26B1, DDX19B, FGF7, PBRM1, PHACTR2, RBM3, RGS7BP, and TEAD1. Prediction of target sequences in the HPV 16 genome identified four target sequences for HPV16-mir-H1-1, two of which are in the E5 gene and two in the L1 gene. HPV16-mir-H2-1 had two targets in the viral genome; one in the LCR region and the other located in the L1 gene.

5. Discussion

Viruses are parasites and thus dependent on the host cellular machinery to complete their life cycle. Therefore, complex relationships have evolved between viruses and the host cells. The role of miRNAs has added yet another dimension to the complexity of viral host interactions.

In Paper I, the effect of HPV16 E5 oncoprotein on the expression of cellular protein-coding genes and miRNA in HaCaT epithelial cell was investigated in genome-wide microarray experiments. Among cellular miRNA, we further validated and investigated the altered expression of hsa-mir-146a (constantly induced by E5), hsa-mir-324-5p (constantly repressed by E5), and hsa-mir-203 (repressed at late time points). Functional annotation analysis of targets of these miRNAs revealed that they were enriched in cell adhesion, cell junction, cell migration, and cell motility. The same dexamethasone-inducible promoter was present in both HaCaT-E5 and HaCaT-E5 cell lines and both cell lines were induced identically, so the effect of dexamethasone should theoretically be subtracted when you compare the two cell lines. Nothing is expressed from the empty vector in control cells by dexamethasone induction but other possible effects of dexamethasone might appear in both cell lines.

Strong upregulation of hsa-mir-146a has been continuously shown during the time frame of 0-96h in our study. This microRNA has been previously investigated in the regulation of immune responses in epithelial disorders (Sonkoly *et al.* 2008). It has been found to be upregulated in breast, pancreatic, and prostate cancers (Volinia *et al.* 2006; Lajer *et al.* 2012), and downregulated in androgen-dependent prostate cancers (Lin *et al.* 2008). It has also been found to be upregulated in cervical cancers (Lajer *et al.* 2012) and to promote cell proliferation when introduced into cervical epithelial cell lines (Wang *et al.* 2008). Hsa-mir-146 expression is induced by ligands of a subset of toll-like receptors (TLR) recognizing bacterial antigens, as well as by TNF-α and IL-1β in a NFκB-dependent manner (Taganov *et al.* 2006). Indeed, we were able to confirm that TNF-α stimulation in E5 cells where hsa-mir-146a expression had been inhibited resulted in considerable activation of the downstream kinase p38 and modest activation of ERK1/2 (Table 3 in Paper I). This shows that upregulation of hsa-mir-146a by E5 may play a significant role in attenuated immune response in HPV infections.

A less studied species, hsa-mir-324-5p, was identified as constantly repressed in E5-expressing cells. Another study also confirmed its repression in HPV positive samples (Lajer *et al.* 2012). Hsa-mir-324-5p is a negative regulator of the oncogenic Hedgehog pathway in neuronal tumors (Ferretti *et al.* 2008). Among the putative hsa-mir-324-5p targets, we showed strong upregulation of N-Cadherin gene and protein expression (Table 3, Figure 2 &3 in Paper I), in agreement with downregulation of hsa-mir-324-5p. Expression of another putative target of hsa-mir-324-5p, E-Cadherin, was increased at the protein level (Figure 2 & 3 in Paper I). Our data indicates that the HPV E5 oncogene

may repress hsa-mir-324-5p expression in cervical epithelial cells and thus contribute to the carcinogenic process. Altogether, our results suggest an involvement of hsa-mir-324-5p in the oncogenic functions of E5.

Hsa-mir-203 was repressed in the later time points. Hsa-mir-203 was the first identified epithelium and skin specific miRNA (Sonkoly *et al.* 2007). It has tumor suppressor function and its downregulation has been observed in tumors (Furuta *et al.* 2010). A recent study proved that HPV E7 downregulates hsa-mir-203 (Melar-New *et al.* 2010). We observed a slight induction of hsa-mir-203 target p63 in E5 induced cells. We further confirmed this by showing that p63 was abolished upon overexpression of hsa-mir-203 (Figure 6 in Paper I). p63 is an important family of transcription factors that maintain the proliferation of basal epithelial cells, and the expression of p63 is a reduced correlate of E5 expression. This finding suggests a connection between HPV oncogene expression and diminished differentiation.

Genome-wide expression data from microarrays or NGS, have been exploited to unravel the complex regulatory networks by miRNAs at a system level (Licatalosi *et al.* 2010). In Paper I, the integrative computational method used was based on the miRNA-mRNA correlation matrix in a time series expression profiles of miRNAs and mRNAs. The predicted miRNA targets have played a critical role in identifying the biological effects of miRNAs (Calin *et al.* 2006). Eight computational tools have been developed for the effective prediction of miRNA target sites in mRNAs (Paper I). We utilized several of these software in concert to identify miRNA functions through their targets, as described in Paper I. However, to reveal the regulatory networks by miRNAs, more experimental data is needed in addition to more elaborate statistical approaches. Single gene expression regulation is subject to various levels of control and may involve a mixture of multiple TFs and miRNAs, which make the finding difficult. The topological structures of regulatory networks include a hierarchical structure and motifs enrichment, such as feed-forward loops (Cheng *et al.* 2011).

New results from high-throughput profiling of small RNAs using NGS technologies have given more supports to the diversity of miRNA variants, isomiRs (Lee *et al.* 2010). IsomiRs can be generated by several mechanisms such as differential processing by Dicer, or RNA editing. Though many miRNA variants are usually expressed in a low level compared to the mature miRNA, some isomiRs are relatively highly represented, differentially expressed, and thus might have biological functions (Morin *et al.* 2008). The isomiRs with 5′ end variants are more critical because the miRNA seed region will also be changed in these isomiRs.

The first step in understanding the miRNA–mRNA regulations would be to find real miRNA targets, which is still a challenge from the perspective of the bioinformatics algorithms development. Currently, the widely used methods for miRNA target prediction rely on the seed sequence complementarity and

conservation across species. However, they generally have high false-positive rates and the results are not consistent among different methods. Some new statistical models and algorithms, such as Bayesian algorithm (Huang *et al.* 2007), classification based algorithm (Hecker *et al.* 2013), and LASSO regression algorithm (Muniategui *et al.* 2012), try to combine the expression data generated from microarray or NGS to predict reliable connections between miRNA-mRNA pairs. Still, these methods cannot distinguish direct and indirect regulation by miRNAs.

The development of NGS technologies have made it possible to identify miRNA target binding sites directly on a global scale by crosslinking and immunoprecipitation high-throughput sequencing (CLIP-Seq or HITS-CLIP) (Licatalosi *et al.* 2008). It utilizes UV-crosslinking between RNA and the protein, followed by sequencing the fragmentation of the crosslinked and coimmunoprecipitated RNA with the protein antibodies. Improved protocols also increased resolution by using photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation, (PAR-CLIP) (Hafner *et al.* 2010) and individual nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP) (Konig *et al.* 2010). PAR-CLIP employs photoreactive ribonucleoside analogs incorporated into nascent RNA transcripts to induce efficient crosslinking with RBPs and miRNPs. The iCLIP relies on the same crosslinking and immunoprecipitation methods as CLIP-Seq. When the RNA is reverse transcribed, iCLIP add 5' adaptor to truncate at the crosslink site. This step allows identification of RNA-protein interaction sites at high resolution and low error rate. New bioinformatics approaches have been developed to analyze these kinds of data (Chou *et al.* 2013; Majoros *et al.* 2013). Using these technologies, new findings are constantly added to the miRNA target knowledgebase. Interestingly, miRNAs bind more often to coding exons than previously thought, and many binding sites seem to deviate from the known target prediction rules (Majoros *et al.* 2013). Combining CLIP-seq results with miRNA/gene expression results from microarray or miRNA/RNA-seq will give more robust result of miRNA targets and the about the functionality of the identified binding sites.

MiRNA characteristics make them attractive for viruses to regulate gene expression. Since they are small in size, they can easily be fitted in the small size genomes. Moreover, miRNAs are non-immunogenic and they take advantage of the host regulatory mechanism. Furthermore, they can target both virus-encoded mRNAs and host gene transcripts. Many dsDNA viruses, such as polyomaviruses, encode miRNAs (Sullivan *et al.* 2005; Seo *et al.* 2008; Chen *et al.* 2011). Papillomaviruses have been suspected to encode their own miRNA because they have dsDNA genomes, they replicate mainly in the nucleus, and they have the ability to establish persistent infection and latency (Cullen 2006). However until now no papillomavirus miRNAs have been validated. Utilizing miRNA sequencing technology will give the advantages of identifying novel papillomavirus encoded miRNA over other traditional

methods.

Since the early days of miRNA sequencing, many bioinformatics tools have been developed to identify miRNAs from millions or even billions of sequencing reads. Several of them have been compared and reviewed in (Li *et al.* 2012) and (Williamson *et al.* 2013), including miRDeep (Friedlander *et al.* 2008), miRDeep2 (Friedlander *et al.* 2012), miRanalyzer (Hackenberg *et al.* 2009), miRExpress (Wang *et al.* 2009), miRTRAP (Hendrix *et al.* 2010), DSAP (Huang *et al.* 2010), mirTools (Zhu *et al.* 2010), MIReNA (Mathelier *et al.* 2010), miRNAkey (Ronen *et al.* 2010), and mireap (Li *et al.* 2012). Our new analysis workflow for miRNA sequencing data processing in R environment was named as miRSeqNovel. Unlike many available tools, miRSeqNovel aims at being a universal workflow instead of being data source specific. It supports results from different mapping methods with different sets of parameters. It has no limitation to what reference genome(s) are used. This is the first miRNA-seq analysis workflow implemented in R/Bioconductor, and it is compatible with all operating systems. When researchers select the optimal methods for miRNA-seq analysis, it is always recommended to understand and verify the software limitations and their specific requirements.

From the miRNA-seq data, the sequencing reads number can easily quantify expression levels of known miRNAs with annotation in miRBase. The benefit of sequencing is the direct detection of isomiR expression. miRSeqNovel considers both annotated mature miRNA and isomiRs in differentially expressed miRNA findings. There are already widely used high-throughput platforms for miRNA analysis, Agilent/Affymetrix microarrays and NGS. It is important and very interesting to compare results between different microarray and NGS platforms. Every platform has its strengths and weaknesses, which might also lead to biases in the result. By using mature spiked-in miRNAs as a positive marker, people found Agilent was the most accurate platform (Leshkowitz *et al.* 2013). However, this result may be improved by applying future statistical methods and may be dependent on the spiked-in miRNAs sequences.

Because miRNA biogenesis involves different mechanisms in different species, it is difficult to define a standard set of parameters for predicting novel miRNAs. miRSeqNovel uses fully customizable parameters on prediction of novel miRNA candidates. Moreover, it predicts conservative novel miRNA candidates without annotation of known miRNAs, which is useful for species with no miRNA annotations. Various computational analyses rely on machine learning approaches to identify novel miRNAs by utilizing all the information publically available in miRBase (Batuwita *et al.* 2009; Kadri *et al.* 2009; van der Burgt *et al.* 2009). However, annotation information in miRBase is updated biannually and still needs further experimental validation (Hansen *et al.* 2011). These methods rarely update their training data and few of them consider species-specific records.

In miRNA sequencing data, reads are usually mapped to hundreds of thousands of genomic locations. However, with a limited number of known miRNAs, it is still a challenge to capture novel miRNAs *in silico* (Chung *et al.* 2011). When comparing the novel miRNA prediction result from available miRNA-seq software, the challenge is that methods are sensitive to the species selected, which can affect the performance. The overlap between individual prediction methods is much less than if combined, implying a high number of false positives (Li *et al.* 2012). The reason for this is that current miRNA sequencing methodology focuses solely on the mature sequence and the precursor is generally a theoretical prediction based on the information provided by the mapped read. Consequently, results can also be affected by the methods and parameters used for mapping (Williamson *et al.* 2013). In order to resolve the apparent discrepancy in predictions of novel miRNA, one needs better methods, both computational and experimental. Unlike previously published methods, miRSeqNovel performs conservative prediction and usually produces a higher number of potential miRNA candidates than other predicting methods, which use mathematical models to filter candidates. The prediction results of known miRNAs indicate that miRSeqNovel can give reliable results (around 80%) without further filtering based on comparing published sets of data.

Screening for candidate HPV miRNAs, miRSeqNovel used current understanding of human miRNA and small ds-virus miRNA features. Accumulating miRNA sequencing data continuously corrects miRNA annotations in the miRBase (Kozomara *et al.* 2011). We considered the relative sequence abundance as one of the main criteria in prediction of mature miRNAs (Kuchenbauer *et al.* 2008). Predicting the precise features of novel miRNAs and star miRNA/isomiRs is challenging, as many candidates had low read counts. Solving this in the second prediction round, we combined all the libraries together using all sequencing reads. Comparison of the results of the first and second prediction rounds revealed advantages and disadvantages in both rounds. The second round will give a better prediction of pre-miRNA when star miRNA, is not fully covered by the pre-miRNA in the first round. However, in the second round, some low expression reads noise may add additional nucleotides to pre-miRNA leading to suboptimal RNA secondary structure such as HPV16-mir-H2. HPV16-mir-H1 was shown using both prediction runs.

Viral miRNAs may also have non-canonical features that differentiate them from human miRNAs (Reese *et al.* 2010). Due to the low expression level of HPV16-mir-H1, it was not possible to establish its exact length. The precursor sequence of HPV16-mir-H1 is still a prediction and hence has uncertainties. This is similar to the known Merkel cell polyomavirus encoded MCV-mir-M1-5p, which was first predicted from VMir followed by validation (Seo *et al.* 2009). It was further identified by Illumina sequencing and validated by qRT-PCR. It exhibits a 5'end 2-nt shift from the VMir predicted MCV-mir-M1 mature

sequence (Lee *et al.* 2011). Further studies are needed to prove whether the isomiRs presented here could also exist and be functional under some conditions.

Viral miRNAs are ideal tools for viruses for modulating both viral and cellular gene expression. Viral miRNAs of SV40 (Sullivan *et al.* 2005), Merkel cell polyomavirus (MCV) (Seo *et al.* 2009; Lee *et al.* 2011), JC virus, BK virus (Seo *et al.* 2008), and BPCV (Chen *et al.* 2011) share similar functions in negatively regulating viral early gene expression by targeting early transcripts (T antigen) with subsequent escape from host immune attack and facilitation of viral replication. Despite the lack of sequence similarity, HPV has similar genome size and similar gene functions to those of polyomaviruses, which suggests that HPV might encode miRNA with related functions.

Entire tissue samples were used for these studies, while only a subset of these cells were infected. This caused low counts of sequencing reads for viral miRNA prediction, which can be improved by deeper sequencing. Robust signals of viral miRNAs were seen in cervical tissue in *in situ* hybridization, often colocalizing with and restricted to p16INK4a signal. *In situ* hybridization also showed altered distribution of human miR-205, which has been shown to promote proliferation of human cervical cancer cells (Xie *et al.* 2012). However, the signals in the *in situ* assays for the U6, miR-205, and HPV miRNAs cannot be directly compared as a measure of the expression level because the signal intensity also depends on the antibodies.

Cellular targets of HPV-encoded miRNAs give an overview of their putative functions in regulating cellular gene expression. Gene ontology analysis of predicted cellular targets of HPV16-mir-H1-1 suggests important roles in host cell interactions of HPV16.This includes the cell cycle process, especially the M phase, which is important for viral infection (Pyeon *et al.* 2009). A set of predicted target genes are involved in regulation of immune functions of the host, such as T cell activation, and immune system development. Also, genes involved in epithelium development and cancer point to an auxiliary role of this miRNA in the onset of epithelial abnormalities and oncogenesis. It also shares the same target gene predicted for Merkel cell polyomavirus miRNA, PIK3CD, which is involved in activation of cell growth, survival, proliferation, and motility, in regulation of cell morphology, and in mediating host immune responses (Darmochwal-Kolarz *et al.* 2009). Further, HPV16-mir-H1-1 has two putative target sites within the E5 gene, suggesting that HPV-encoded miRNAs may be involved in activation of the Akt/PI3K pathway by the HPV 16 E5 oncogene (Kivi *et al.* 2008).

HPV16-mir-H2-1 predicted cellular targets are also presented in similar pathways, like cell cycle process and M phase, and in immune regulation such as T cell activation and immune system development. Importantly, predicted target gene functions in cell migration and cell adhesion showed very similar functions in earlier studies (Kivi *et al.* 2008), suggesting a possible involvement

or interplay with the E5 oncogene.

Expression levels of HPV-encoded miRNAs described here were low, which is reasonable given that even low levels may suffice to facilitate viral replication, and that their targets may also be important for viral replication. The predicted viral microRNA also target sites within the E5 gene, L1 gene, or LCR in the viral genome (Figure 1 in Paper III). Autoregulation of viral replication as shown for polyomavirus miRNA establishes latency (Skalsky *et al.* 2010). E5 transcripts of genital papillomaviruses are always multicistronic (Wang *et al.* 2011). How the HPV miRNAs target these particular regions would affect the expression of several viral genes and increase the possibility of pathogenesis of the papillomaviruses.

6. Conclusions

Both microarray and NGS methods are suitable to study the genome-wide gene expression profiling. We believe that HPV 16 E5 oncogene contributes to carcinogenesis by several mechanisms that involve regulation of cellular miRNA and their target genes. Alterations in miRNA expression of HPV 16 E5 oncogene seem to increase cell proliferation and tumorigenesis, and repress epithelial differentiation. As compared to microarray, NGS has technical advantages but there is still a need for better bioinformatics solutions. The miRNA-Seq has become a very popular tool to reveal the miRNA expression profiles and find novel miRNAs. miRSeqNovel is a fully customizable method implemented in R. It is flexible and easy to use given that the user is familiar with miRNA biology and R/Bioconductor basics. It accepts both colorspace (SOLiD) and basespace (Illumina/Solexa) data. It finds differentially expressed miRNAs and gives conservative prediction of novel miRNA candidates with customized parameters. We discovered novel HPV-encoded miRNAs by SOLiD 4 and further validated them by qRT-PCR and *in situ* hybridization. Biological functions of the predicted cellular target genes suggest a possible and significant role in the establishment of infection and in carcinogenesis. Viral miRNA are also tempting as possible targets for new antiviral drugs. These findings emphasize the need for further studies on HPV miRNA functions.

7. Acknowledgements

This work has been carried out in the DNA Sequencing and Genomics Laboratory in the Institute of Biotechnology, University of Helsinki, with financial support from the University of Helsinki and Viikki Doctoral Programme in Molecular Biosciences (VGSB).

I would like to express my profound gratitude to my thesis supervisors, Docent Petri Auvinen and Docent Dario Greco. From you, I learnt one person can really have such diverse and deeper understanding of multiple disciplines. You are my standard of a great scientist. Your way of thinking and handling scientific questions inspires me a lot. Moreover, the relaxed atmosphere in the group and your patient advising and your faith in me has made me really enjoy the whole process. I appreciate all the scientific and non-scientific discussions we had.

This thesis would not be possible without the contribution, support and help from other excellent people. First, I would like to thank to Docent Eeva Auvinen, one member of my follow up group and a great collaborator, for helping me a lot in advising and supporting my work, and improving my writing skills. Second, I would like to thank all collaborators, Dr. Nina Kivi, M.Sc Tuuli Pietilä, Dr. Mikko Rönty, Dr. Suvi-Katri Leivonen, Dr. Mikko Frilander, Dr. Frederic Michon, Dr. Jarmo Ritari, Dr. Jussi Tarkkanen, Dr. Lars Paulín. All of you have expert skills in your area and thank you for your contribution to these projects. Finally, I want to thank Mrs Eeva-Marja Turkki and Mrs Kirsi Lipponen for performing the SOLiD sequencing in our lab.

I would also like to thank Professor Jukka Corander, also my follow up group member, for the wonderful advice and support, especially for mathematical solutions. A special thanks to Docent Janna Saarela and Professor Veijo Hukkanen for reviewing my thesis. I wish to thank Prof. Tapio Palva and Dr. Pekka Heino at the Department of Biosciences in Faculty of Biological and Environmental Sciences. Dr. Dennis Bamford, Dr. Sandra Falck and Dr. Ulla Tuomainen of the Viikki Doctoral Programme in Molecular Biosciences (VGSB) are also acknowledged. I want to thank Professor Mart Saarma, former Director of the Institute of Biotechnology, and Professor Tomi P. Mäkelä, present Director of the Institute of Biotechnology, for giving us all an excellent work environment and facilities in the Institute.

I am grateful for all lab members during these years. It was always nice to celebrate birthdays or have some recreations together. Also special thanks to Dr. Panu Somervuo and Olli Lounela, for wonderful discussions on bioinformatics and computers.

I also wish to thank all my Chinese friends here. I thank Feng Lin and Zeng Zao for their help when I just arrived in Finland. The hours I spent playing Wii with Cheng Jianping and Lan Qiang, and Xbox with Yang Ying and Tan Miaoqing were great fun. My colorful life in Helsinki would not have been possible without happy times with Zheng Wei, Gao Yajing, Yan Yan, Fang Shentong, Cheng Fang, Yan Jian, Cheng Lu, Zhou You, Yin Miao, Wei Jing, Hu Yizhou, Cheng Ping, Li Songping, Wei Hong, Jiang Ping, Wei Gonghong, Wu Jianmin, Li Chunmei, Tian Li, Ma Li, Li Zhilin, Wang Hao, Gui Jinghua and many other friends.

I also thank my parents for a lifetime of support, understanding and encouragement. I thank all my relatives who care about my study, my life and my future. I also want to thank several of my best friends all over the world: Jia Lei, Teng Teng, Li Jingyi for our forever friendship.

Thank you.

Kui Qian (钱夔)
Helsinki, September 2013

8. References

Aboyoun P, Pages H and Lawrence M (2013). "GenomicRanges: Representation and manipulation of genomic intervals." R package version 1.12.5.

Anders S and Huber W (2010). "Differential expression analysis for sequence count data." Genome Biol 11(10): R106.

Ashburner M, Ball C A, Blake J A, Botstein D, Butler H, Cherry J M, Davis A P, Dolinski K, Dwight S S*, et al.* (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." Nat Genet 25(1): 25-29.

Au Yeung C L, Tsang T Y, Yau P L and Kwok T T (2011). "Human papillomavirus type 16 E6 induces cervical cancer cell migration through the p53/microRNA-23b/urokinase-type plasminogen activator pathway." Oncogene 30(21): 2401-2410.

Balsitis S, Dick F, Dyson N and Lambert P F (2006). "Critical roles for non-pRb targets of human papillomavirus type 16 E7 in cervical carcinogenesis." Cancer Res 66(19): 9393-9400.

Batuwita R and Palade V (2009). "microPred: effective classification of pre-miRNAs for human miRNA gene prediction." Bioinformatics 25(8): 989-995.

Bauman Y, Nachmani D, Vitenshtein A, Tsukerman P, Drayman N, Stern-Ginossar N, Lankry D, Gruda R and Mandelboim O (2011). "An identical miRNA of the human JC and BK polyoma viruses targets the stress-induced ligand ULBP3 to escape immune elimination." Cell Host Microbe 9(2): 93-102.

Benjamini Y and Hochberg Y (1995). "Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing." Journal of the Royal Statistical Society Series B-Methodological 57(1): 289-300.

Bentley D R, Balasubramanian S, Swerdlow H P, Smith G P, Milton J, Brown C G, Hall K P, Evers D J, Barnes C L*, et al.* (2008). "Accurate whole human genome sequencing using reversible terminator chemistry." Nature 456(7218): 53-59.

Berezikov E, Chung W J, Willis J, Cuppen E and Lai E C (2007). "Mammalian mirtron genes." Mol Cell 28(2): 328-336.

Bernard H U, Burk R D, Chen Z, van Doorslaer K, Hausen H and de Villiers E M (2010). "Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments." Virology 401(1): 70-79.

Betel D, Wilson M, Gabow A, Marks D S and Sander C (2008). "The microRNA.org resource: targets and expression." Nucleic Acids Res 36(Database issue): D149-153.

Bodily J and Laimins L A (2011). "Persistence of human papillomavirus infection: keys to malignant progression." Trends Microbiol 19(1): 33-39.

Borralho P M, Simoes A E, Gomes S E, Lima R T, Carvalho T, Ferreira D M, Vasconcelos M H, Castro R E and Rodrigues C M (2011). "miR-143 overexpression impairs growth of human colon carcinoma xenografts in mice with induction of apoptosis and inhibition of proliferation." PLoS One 6(8): e23787.

Cai X, Hagedorn C H and Cullen B R (2004). "Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs." RNA 10(12): 1957-1966.

Cai X, Li G, Laimins L A and Cullen B R (2006). "Human papillomavirus genotype 31 does not express detectable microRNA levels during latent or productive virus replication." J Virol 80(21): 10890-10893.

Calin G A and Croce C M (2006). "MicroRNA signatures in human cancers." Nat Rev Cancer 6(11): 857-866.

Chaturvedi A K, Katki H A, Hildesheim A, Rodriguez A C, Quint W, Schiffman M, Van Doorn L J, Porras C, Wacholder S*, et al.* (2011). "Human papillomavirus infection with multiple types: pattern of coinfection and risk of cervical disease." J Infect Dis 203(7): 910-920.

Chen C J, Kincaid R P, Seo G J, Bennett M D and Sullivan C S (2011). "Insights into Polyomaviridae microRNA function derived from study of the bandicoot papillomatosis carcinomatosis viruses." J Virol 85(9): 4487-4500.

Cheng C, Yan K K, Hwang W, Qian J, Bhardwaj N, Rozowsky J, Lu Z J, Niu W, Alves P*, et al.* (2011). "Construction and analysis of an integrated regulatory network derived from high-throughput sequencing data." PLoS Comput Biol 7(11): e1002190.

Chou C H, Lin F M, Chou M T, Hsu S D, Chang T H, Weng S L, Shrestha S, Hsiao C C, Hung J H*, et al.* (2013). "A computational approach for identifying microRNA-target interactions using high-throughput CLIP and PAR-CLIP sequencing." BMC Genomics 14 Suppl 1: S2.

Chow L T, Broker T R and Steinberg B M (2010). "The natural history of human papillomavirus infections of the mucosal epithelia." APMIS 118(6-7): 422-449.

Chung W J, Agius P, Westholm J O, Chen M, Okamura K, Robine N, Leslie C S and Lai E C (2011). "Computational and experimental identification of mirtrons in Drosophila melanogaster and Caenorhabditis elegans." Genome Res 21(2): 286-300.

Cloonan N, Wani S, Xu Q, Gu J, Lea K, Heater S, Barbacioru C, Steptoe A L, Martin H C*, et al.* (2011). "MicroRNAs and their isomiRs function cooperatively to target common biological pathways." Genome Biol 12(12): R126.

Coupe V M, González-Barreiro L, Gutiérrez-Berzal J, Melián-Boveda A L, López-Rodríguez O, Alba-Domínguez J and Alba-Losada J (2012). "Transcriptional analysis of human papillomavirus type 16 in histological sections of cervical dysplasia by *in situ* hybridisation." J Clin Pathol

65(2): 164-170.
Crusius K, Auvinen E and Alonso A (1997). "Enhancement of EGF- and PMA-mediated MAP kinase activation in cells expressing the human papillomavirus type 16 E5 protein." Oncogene 15(12): 1437-1444.
Crusius K, Auvinen E, Steuer B, Gaissert H and Alonso A (1998). "The human papillomavirus type 16 E5-protein modulates ligand-dependent activation of the EGF receptor family in the human epithelial cell line HaCaT." Exp Cell Res 241(1): 76-83.
Cullen B R (2006). "Viruses and microRNAs." Nat Genet 38 Suppl: S25-30.
Darmochwal-Kolarz D, Gasowska-Giszczak U, Paduch R, Kolarz B, Wilcinski P, Oleszczuk J and Kwasniewska A (2009). "Apoptosis of HeLa and CaSki cell lines incubated with All-trans retinoid acid." Folia Histochem Cytobiol 47(4): 599-603.
Dennis G, Jr., Sherman B T, Hosack D A, Yang J, Gao W, Lane H C and Lempicki R A (2003). "DAVID: Database for Annotation, Visualization, and Integrated Discovery." Genome Biol 4(5): P3.
Dillies M A, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D*, et al.* (2012). "A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis." Brief Bioinform.
DiMaio D and Mattoon D (2001). "Mechanisms of cell transformation by papillomavirus E5 proteins." Oncogene 20(54): 7866-7873.
Doorbar J (2006). "Molecular biology of human papillomavirus infection and cervical cancer." Clin Sci (Lond) 110(5): 525-541.
Doorbar J, Foo C, Coleman N, Medcalf L, Hartley O, Prospero T, Napthine S, Sterling J, Winter G*, et al.* (1997). "Characterization of events during the late stages of HPV16 infection in vivo using high-affinity synthetic Fabs to E4." Virology 238(1): 40-52.
Doorbar J, Quint W, Banks L, Bravo I G, Stoler M, Broker T R and Stanley M A (2012). "The biology and life-cycle of human papillomaviruses." Vaccine 30 Suppl 5: F55-70.
Duggan D J, Bittner M, Chen Y, Meltzer P and Trent J M (1999). "Expression profiling using cDNA microarrays." Nat Genet 21(1 Suppl): 10-14.
Dürst M, Dzarlieva-Petrusevska R T, Boukamp P, Fusenig N E and Gissmann L (1987). "Molecular and cytogenetic analysis of immortalized human primary keratinocytes obtained after transfection with human papillomavirus type 16 DNA." Oncogene 1(3): 251-256.
Dweep H, Sticht C, Pandey P and Gretz N (2011). "miRWalk - Database: Prediction of possible miRNA binding sites by "walking" the genes of three genomes." J Biomed Inform.
Dyson N, Howley P M, Münger K and Harlow E (1989). "The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product." Science 243(4893): 934-937.
Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P*,*

et al. (2009). "Real-time DNA sequencing from single polymerase molecules." Science 323(5910): 133-138.
Esquela-Kerscher A and Slack F J (2006). "Oncomirs - microRNAs with a role in cancer." Nat Rev Cancer 6(4): 259-269.
Fan J B, Chee M S and Gunderson K L (2006). "Highly parallel genomic assays." Nat Rev Genet 7(8): 632-644.
Feng Z, Fang G, Korlach J, Clark T, Luong K, Zhang X, Wong W and Schadt E (2013). "Detecting DNA modifications from SMRT sequencing data by modeling sequence context dependence of polymerase kinetic." PLoS Comput Biol 9(3): e1002935.
Ferretti E, De Smaele E, Miele E, Laneve P, Po A, Pelloni M, Paganelli A, Di Marcotullio L, Caffarelli E*, et al.* (2008). "Concerted microRNA control of Hedgehog signalling in cerebellar neuronal progenitor and tumour cells." EMBO J 27(19): 2616-2627.
Flicek P, Amode M R, Barrell D, Beal K, Brent S, Chen Y, Clapham P, Coates G, Fairley S*, et al.* (2011). "Ensembl 2011." Nucleic Acids Res 39(Database issue): D800-806.
Friedlander M R, Chen W, Adamidi C, Maaskola J, Einspanier R, Knespel S and Rajewsky N (2008). "Discovering microRNAs from deep sequencing data using miRDeep." Nat Biotechnol 26(4): 407-415.
Friedlander M R, Mackowiak S D, Li N, Chen W and Rajewsky N (2012). "miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades." Nucleic Acids Res 40(1): 37-52.
Furuta M, Kozaki K I, Tanaka S, Arii S, Imoto I and Inazawa J (2010). "miR-124 and miR-203 are epigenetically silenced tumor-suppressive microRNAs in hepatocellular carcinoma." Carcinogenesis 31(5): 766-776.
Gao P and Zheng J (2010). "High-risk HPV E5-induced cell fusion: a critical initiating event in the early stage of HPV-associated cervical cancer." Virol J 7: 238.
Gebeshuber C A, Zatloukal K and Martinez J (2009). "miR-29a suppresses tristetraprolin, which is a regulator of epithelial polarity and metastasis." EMBO Rep 10(4): 400-405.
Genther Williams S M, Disbrow G L, Schlegel R, Lee D, Threadgill D W and Lambert P F (2005). "Requirement of epidermal growth factor receptor for hyperplasia induced by E5, a high-risk human papillomavirus oncogene." Cancer Res 65(15): 6534-6542.
Gerstein M B, Kundaje A, Hariharan M, Landt S G, Yan K K, Cheng C, Mu X J, Khurana E, Rozowsky J*, et al.* (2012). "Architecture of the human regulatory network derived from ENCODE data." Nature 489(7414): 91-100.
Gocze K, Gombos K, Juhasz K, Kovacs K, Kajtar B, Benczik M, Gocze P, Patczai B, Arany I*, et al.* (2013). "Unique microRNA expression profiles

in cervical cancer." Anticancer Res 33(6): 2561-2567.
Grimson A, Farh K K, 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(1): 91-105.
Grundhoff A and Sullivan C S (2011). "Virus-encoded microRNAs." Virology 411(2): 325-343.
Gu W, An J, Ye P, Zhao K N and Antonsson A (2011). "Prediction of conserved microRNAs from skin and mucosal human papillomaviruses." Arch Virol 156(7): 1161-1171.
Guan P, Howell-Jones R, Li N, Bruni L, de Sanjosé S, Franceschi S and Clifford G M (2012). "Human papillomavirus types in 115,789 HPV-positive women: A meta-analysis from cervical infection to cancer." Int J Cancer 131(10): 2349-2359.
Hackenberg M, Sturm M, Langenberger D, Falcon-Perez J M and Aransay A M (2009). "miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments." Nucleic Acids Res 37(Web Server issue): W68-76.
Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jr., Jungkamp A C*, et al.* (2010). "Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP." Cell 141(1): 129-141.
Hansen T B, Kjems J and Bramsen J B (2011). "Enhancing miRNA annotation confidence in miRBase by continuous cross dataset analysis." RNA Biol 8(3): 378-383.
Hawkins R D, Hon G C and Ren B (2010). "Next-generation genomics: an integrative approach." Nat Rev Genet 11(7): 476-486.
Hayashita Y, Osada H, Tatematsu Y, Yamada H, Yanagisawa K, Tomida S, Yatabe Y, Kawahara K, Sekido Y*, et al.* (2005). "A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation." Cancer Res 65(21): 9628-9632.
He L, Thomson J M, Hemann M T, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe S W*, et al.* (2005). "A microRNA polycistron as a potential human oncogene." Nature 435(7043): 828-833.
Heck D V, Yee C L, Howley P M and Münger K (1992). "Efficiency of binding the retinoblastoma protein correlates with the transforming capacity of the E7 oncoproteins of the human papillomaviruses." Proc Natl Acad Sci U S A 89(10): 4442-4446.
Hecker N, Stephan C, Mollenkopf H J, Jung K, Preissner R and Meyer H A (2013). "A New Algorithm for Integrated Analysis of miRNA-mRNA Interactions Based on Individual Classification Reveals Insights into Bladder Cancer." PLoS One 8(5): e64543.
Hendrix D, Levine M and Shi W (2010). "miRTRAP, a computational method for the systematic identification of miRNAs from high throughput

sequencing data." Genome Biol 11(4): R39.
Hofacker I L (2003). "Vienna RNA secondary structure server." Nucleic Acids Res 31(13): 3429-3431.
Hsu S D, Lin F M, Wu W Y, Liang C, Huang W C, Chan W L, Tsai W T, Chen G Z, Lee C J*, et al.* (2011). "miRTarBase: a database curates experimentally validated microRNA-target interactions." Nucleic Acids Res 39(Database issue): D163-169.
Hu X, Schwarz J K, Lewis J S, Jr., Huettner P C, Rader J S, Deasy J O, Grigsby P W and Wang X (2010). "A microRNA expression signature for cervical cancer prognosis." Cancer Res 70(4): 1441-1448.
Huang da W, Sherman B T and Lempicki R A (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nat Protoc 4(1): 44-57.
Huang J C, Babak T, Corson T W, Chua G, Khan S, Gallie B L, Hughes T R, Blencowe B J, Frey B J*, et al.* (2007). "Using expression profiling data to identify human microRNA targets." Nat Methods 4(12): 1045-1049.
Huang P J, Liu Y C, Lee C C, Lin W C, Gan R R, Lyu P C and Tang P (2010). "DSAP: deep-sequencing small RNA analysis pipeline." Nucleic Acids Res 38(Web Server issue): W385-391.
Hughes T R, Mao M, Jones A R, Burchard J, Marton M J, Shannon K W, Lefkowitz S M, Ziman M, Schelter J M*, et al.* (2001). "Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer." Nat Biotechnol 19(4): 342-347.
Hussain M, Taft R J and Asgari S (2008). "An insect virus-encoded microRNA regulates viral replication." J Virol 82(18): 9164-9170.
Iorio M V, Ferracin M, Liu C G, Veronese A, Spizzo R, Sabbioni S, Magri E, Pedriali M, Fabbri M*, et al.* (2005). "MicroRNA gene expression deregulation in human breast cancer." Cancer Res 65(16): 7065-7070.
Iorio M V, Visone R, Di Leva G, Donati V, Petrocca F, Casalini P, Taccioli C, Volinia S, Liu C G*, et al.* (2007). "MicroRNA signatures in human ovarian cancer." Cancer Res 67(18): 8699-8707.
Jemal A, Bray F, Center M M, Ferlay J, Ward E and Forman D (2011). "Global cancer statistics." CA Cancer J Clin 61(2): 69-90.
Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, Li M, Wang G and Liu Y (2009). "miR2Disease: a manually curated database for microRNA deregulation in human disease." Nucleic Acids Res 37(Database issue): D98-104.
Jima D D, Zhang J, Jacobs C, Richards K L, Dunphy C H, Choi W W, Au W Y, Srivastava G, Czader M B*, et al.* (2010). "Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs." Blood 116(23): e118-127.
Kadri S, Hinman V and Benos P V (2009). "HHMMiR: efficient de novo prediction of microRNAs using hierarchical hidden Markov models." BMC Bioinformatics 10 Suppl 1: S35.

Kanehisa M, Goto S, Sato Y, Furumichi M and Tanabe M (2012). "KEGG for integration and interpretation of large-scale molecular data sets." Nucleic Acids Res 40(Database issue): D109-114.

Kertesz M, Iovino N, Unnerstall U, Gaul U and Segal E (2007). "The role of site accessibility in microRNA target recognition." Nat Genet 39(10): 1278-1284.

Kim V N, Han J and Siomi M C (2009). "Biogenesis of small RNAs in animals." Nat Rev Mol Cell Biol 10(2): 126-139.

Kim Y K and Kim V N (2007). "Processing of intronic microRNAs." EMBO J 26(3): 775-783.

Kivi N, Greco D, Auvinen P and Auvinen E (2008). "Genes involved in cell adhesion, cell motility and mitogenic signaling are altered due to HPV 16 E5 protein expression." Oncogene 27(18): 2532-2541.

Klaes R, Friedrich T, Spitkovsky D, Ridder R, Rudy W, Petry U, Dallenbach-Hellweg G, Schmidt D and von Knebel Doeberitz M (2001). "Overexpression of p16(INK4A) as a specific marker for dysplastic and neoplastic epithelial cells of the cervix uteri." Int J Cancer 92(2): 276-284.

Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner D J, Luscombe N M and Ule J (2010). "iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution." Nat Struct Mol Biol 17(7): 909-915.

Kozomara A and Griffiths-Jones S (2011). "miRBase: integrating microRNA annotation and deep-sequencing data." Nucleic Acids Res 39(Database issue): D152-157.

Krek A, Grün D, Poy M N, Wolf R, Rosenberg L, Epstein E J, MacMenamin P, da Piedade I, Gunsalus K C*, et al.* (2005). "Combinatorial microRNA target predictions." Nat Genet 37(5): 495-500.

Krol J, Loedige I and Filipowicz W (2010). "The widespread regulation of microRNA biogenesis, function and decay." Nat Rev Genet 11(9): 597-610.

Krueger F, Kreck B, Franke A and Andrews S R (2012). "DNA methylome analysis using short bisulfite sequencing data." Nat Methods 9(2): 145-151.

Kuchenbauer F, Morin R D, Argiropoulos B, Petriv O I, Griffith M, Heuser M, Yung E, Piper J, Delaney A*, et al.* (2008). "In-depth characterization of the microRNA transcriptome in a leukemia progression model." Genome Res 18(11): 1787-1797.

Kumar A, Wong A K, Tizard M L, Moore R J and Lefèvre C (2012). "miRNA_Targets: a database for miRNA target predictions in coding and non-coding regions of mRNAs." Genomics 100(6): 352-356.

Lajer C B, Garnaes E, Friis-Hansen L, Norrild B, Therkildsen M H, Glud M, Rossing M, Lajer H, Svane D*, et al.* (2012). "The role of miRNAs in human papilloma virus (HPV)-associated cancers: bridging between

HPV-related head and neck cancer and cervical cancer." Br J Cancer 106(9): 1526-1534.
Langmead B and Salzberg S L (2012). "Fast gapped-read alignment with Bowtie 2." Nat Methods 9(4): 357-359.
Langmead B, Trapnell C, Pop M and Salzberg S L (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." Genome Biol 10(3): R25.
Lee J W, Choi C H, Choi J J, Park Y A, Kim S J, Hwang S Y, Kim W Y, Kim T J, Lee J H*, et al.* (2008). "Altered MicroRNA expression in cervical carcinomas." Clin Cancer Res 14(9): 2535-2542.
Lee L W, Zhang S, Etheridge A, Ma L, Martin D, Galas D and Wang K (2010). "Complexity of the microRNA repertoire revealed by next-generation sequencing." RNA 16(11): 2170-2180.
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(5): 843-854.
Lee S, Paulson K G, Murchison E P, Afanasiev O K, Alkan C, Leonard J H, Byrd D R, Hannon G J and Nghiem P (2011). "Identification and validation of a novel mature microRNA encoded by the Merkel cell polyomavirus in human Merkel cell carcinomas." J Clin Virol 52(3): 272-275.
Lee Y, Kim M, Han J, Yeom K H, Lee S, Baek S H and Kim V N (2004). "MicroRNA genes are transcribed by RNA polymerase II." EMBO J 23(20): 4051-4060.
Leechanachai P, Banks L, Moreau F and Matlashewski G (1992). "The E5 gene from human papillomavirus type 16 is an oncogene which enhances growth factor-mediated signal transduction to the nucleus." Oncogene 7(1): 19-25.
Leshkowitz D, Horn-Saban S, Parmet Y and Feldmesser E (2013). "Differences in microRNA detection levels are technology and sequence dependent." RNA 19(4): 527-538.
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(1): 15-20.
Lewis B P, Shih I H, Jones-Rhoades M W, Bartel D P and Burge C B (2003). "Prediction of mammalian microRNA targets." Cell 115(7): 787-798.
Li B, Hu Y, Ye F, Li Y, Lv W and Xie X (2010). "Reduced miR-34a expression in normal cervical tissues and cervical lesions with high-risk human papillomavirus infection." Int J Gynecol Cancer 20(4): 597-604.
Li H and Durbin R (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." Bioinformatics 25(14): 1754-1760.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G and Durbin R (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics 25(16): 2078-2079.

Li J B, Levanon E Y, Yoon J K, Aach J, Xie B, Leproust E, Zhang K, Gao Y and Church G M (2009). "Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing." Science 324(5931): 1210-1213.

Li Y, Zhang Z, Liu F, Vongsangnak W, Jing Q and Shen B (2012). "Performance comparison and evaluation of software tools for microRNA deep-sequencing data analysis." Nucleic Acids Res 40(10): 4298-4305.

Licatalosi D D and Darnell R B (2010). "RNA processing and its regulation: global insights into biological networks." Nat Rev Genet 11(1): 75-87.

Licatalosi D D, Mele A, Fak J J, Ule J, Kayikci M, Chi S W, Clark T A, Schweitzer A C, Blume J E*, et al.* (2008). "HITS-CLIP yields genome-wide insights into brain alternative RNA processing." Nature 456(7221): 464-469.

Life Technologies. (2013). "Lifescope software." from http://www.lifetechnologies.com/lifescope.

Lin S L, Chiang A, Chang D and Ying S Y (2008). "Loss of mir-146a function in hormone-refractory prostate cancer." RNA 14(3): 417-424.

Lorenz R, Bernhart S H, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler P F and Hofacker I L (2011). "ViennaRNA Package 2.0." Algorithms Mol Biol 6: 26.

Lorincz A T, Reid R, Jenson A B, Greenberg M D, Lancaster W and Kurman R J (1992). "Human papillomavirus infection of the cervix: relative risk associations of 15 common anogenital types." Obstet Gynecol 79(3): 328-337.

Lui W O, Pourmand N, Patterson B K and Fire A (2007). "Patterns of known and novel small RNAs in human cervical cancer." Cancer Res 67(13): 6031-6043.

Ma D, Zhang Y Y, Guo Y L, Li Z J and Geng L (2012). "Profiling of microRNA-mRNA reveals roles of microRNAs in cervical cancer." Chin Med J (Engl) 125(23): 4270-4276.

Maes O C, Chertkow H M, Wang E and Schipper H M (2009). "MicroRNA: Implications for Alzheimer Disease and other Human CNS Disorders." Curr Genomics 10(3): 154-168.

Majoros W H, Lekprasert P, Mukherjee N, Skalsky R L, Corcoran D L, Cullen B R and Ohler U (2013). "MicroRNA target site identification by integrating sequence and binding information." Nat Methods.

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

Marcet B, Chevalier B, Luxardi G, Coraux C, Zaragosi L E, Cibois M, Robbe-Sermesant K, Jolly T, Cardinaud B*, et al.* (2011). "Control of vertebrate multiciliogenesis by miR-449 through direct repression of the

Delta/Notch pathway." Nat Cell Biol 13(6): 693-699.
Martinez I, Gardiner A S, Board K F, Monzon F A, Edwards R P and Khan S A (2008). "Human papillomavirus type 16 reduces the expression of microRNA-218 in cervical carcinoma cells." Oncogene 27(18): 2575-2582.
Mathelier A and Carbone A (2010). "MIReNA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data." Bioinformatics 26(18): 2226-2234.
Mattick J S (2004). "RNA regulation: a new genetics?" Nat Rev Genet 5(4): 316-323.
Maufort J P, Shai A, Pitot H C and Lambert P F (2010). "A role for HPV16 E5 in cervical carcinogenesis." Cancer Res 70(7): 2924-2931.
Maufort J P, Williams S M, Pitot H C and Lambert P F (2007). "Human papillomavirus 16 E5 oncogene contributes to two stages of skin carcinogenesis." Cancer Res 67(13): 6106-6112.
Mayr C and Bartel D P (2009). "Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells." Cell 138(4): 673-684.
Melar-New M and Laimins L A (2010). "Human papillomaviruses modulate expression of microRNA 203 upon epithelial differentiation to control levels of p63 proteins." J Virol 84(10): 5212-5221.
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(6): 1203-1217.
Moody C A and Laimins L A (2010). "Human papillomavirus oncoproteins: pathways to transformation." Nat Rev Cancer 10(8): 550-560.
Morin R D, O'Connor M D, Griffith M, Kuchenbauer F, Delaney A, Prabhu A L, Zhao Y, McDonald H, Zeng T*, et al.* (2008). "Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells." Genome Res 18(4): 610-621.
Muniategui A, Nogales-Cadenas R, Vazquez M, Aranguren X L, Agirre X, Luttun A, Prosper F, Pascual-Montano A and Rubio A (2012). "Quantification of miRNA-mRNA interactions." PLoS One 7(2): e30766.
Murchison E P and Hannon G J (2004). "miRNAs on the move: miRNA biogenesis and the RNAi machinery." Curr Opin Cell Biol 16(3): 223-229.
Murphy E, Vanicek J, Robins H, Shenk T and Levine A J (2008). "Suppression of immediate-early viral gene expression by herpesvirus-coded microRNAs: implications for latency." Proc Natl Acad Sci U S A 105(14): 5453-5458.
Nadiminty N, Tummala R, Lou W, Zhu Y, Shi X B, Zou J X, Chen H, Zhang J, Chen X*, et al.* (2012). "MicroRNA let-7c is downregulated in prostate cancer and suppresses prostate cancer growth." PLoS One 7(3):

e32832.
Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M and Snyder M (2008). "The transcriptional landscape of the yeast genome defined by RNA sequencing." Science 320(5881): 1344-1349.
Nguyen M, Song S, Liem A, Androphy E, Liu Y and Lambert P F (2002). "A mutant of human papillomavirus type 16 E6 deficient in binding alpha-helix partners displays reduced oncogenic potential in vivo." J Virol 76(24): 13039-13048.
Ozsolak F and Milos P M (2011). "RNA sequencing: advances, challenges and opportunities." Nat Rev Genet 12(2): 87-98.
Pages H, Aboyoun P, Gentleman R and DebRoy S (2013). "Biostrings: String objects representing biological sequences, and matching algorithms." R package version 2.28.0.
Pallante P, Visone R, Ferracin M, Ferraro A, Berlingieri M T, Troncone G, Chiappetta G, Liu C G, Santoro M*, et al.* (2006). "MicroRNA deregulation in human thyroid papillary carcinomas." Endocr Relat Cancer 13(2): 497-508.
Pang R T, Leung C O, Ye T M, Liu W, Chiu P C, Lam K K, Lee K F and Yeung W S (2010). "MicroRNA-34a suppresses invasion through downregulation of Notch1 and Jagged1 in cervical carcinoma and choriocarcinoma cells." Carcinogenesis 31(6): 1037-1044.
Peng R Q, Wan H Y, Li H F, Liu M, Li X and Tang H (2012). "MicroRNA-214 suppresses growth and invasiveness of cervical cancer cells by targeting UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 7." J Biol Chem 287(17): 14301-14309.
Pessa H K, Greco D, Kvist J, Wahlström G, Heino T I, Auvinen P and Frilander M J (2010). "Gene expression profiling of U12-type spliceosome mutant Drosophila reveals widespread changes in metabolic pathways." PLoS One 5(10): e13215.
Pett M and Coleman N (2007). "Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis?" J Pathol 212(4): 356-367.
Pfeffer S and Voinnet O (2006). "Viruses, microRNAs and cancer." Oncogene 25(46): 6211-6219.
Pim D, Collins M and Banks L (1992). "Human papillomavirus type 16 E5 gene stimulates the transforming activity of the epidermal growth factor receptor." Oncogene 7(1): 27-32.
Pop M and Salzberg S L (2008). "Bioinformatics challenges of new sequencing technology." Trends Genet 24(3): 142-149.
Pyeon D, Pearce S M, Lank S M, Ahlquist P and Lambert P F (2009). "Establishment of human papillomavirus infection requires cell cycle progression." PLoS Pathog 5(2): e1000318.
Qian K, Di Lieto A, Corander J, Auvinen P and Greco D (2012). "Re-analysis of bipolar disorder and schizophrenia gene expression complements the

Kraepelinian dichotomy." Adv Exp Med Biol 736: 563-577.
R Core Team (2013). R: A Language and Environment for Statistical Computing.
Reese T A, Xia J, Johnson L S, Zhou X, Zhang W and Virgin H W (2010). "Identification of novel microRNA-like molecules generated from herpesvirus and host tRNA transcripts." J Virol 84(19): 10344-10353.
Ritari J, Hultman J, Fingerroos R, Tarkkanen J, Pullat J, Paulin L, Kivi N, Auvinen P and Auvinen E (2012). "Detection of human papillomaviruses by polymerase chain reaction and ligation reaction on universal microarray." PLoS One 7(3): e34211.
Roberts A P and Jopling C L (2010). "Targeting viral infection by microRNA inhibition." Genome Biol 11(1): 201.
Robinson M D, McCarthy D J and Smyth G K (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." Bioinformatics 26(1): 139-140.
Romanczuk H and Howley P M (1992). "Disruption of either the E1 or the E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity." Proc Natl Acad Sci U S A 89(7): 3159-3163.
Romanczuk H, Thierry F and Howley P M (1990). "Mutational analysis of cis elements involved in E2 modulation of human papillomavirus type 16 P97 and type 18 P105 promoters." J Virol 64(6): 2849-2859.
Ronen R, Gan I, Modai S, Sukacheov A, Dror G, Halperin E and Shomron N (2010). "miRNAkey: a software for microRNA deep sequencing analysis." Bioinformatics 26(20): 2615-2616.
Schena M, Shalon D, Davis R W and Brown P O (1995). "Quantitative monitoring of gene expression patterns with a complementary DNA microarray." Science 270(5235): 467-470.
Schiffman M, Castle P E, Jeronimo J, Rodriguez A C and Wacholder S (2007). "Human papillomavirus and cervical cancer." Lancet 370(9590): 890-907.
Schmitz M, Driesch C, Jansen L, Runnebaum I B and Dürst M (2012). "Non-random integration of the HPV genome in cervical cancer." PLoS One 7(6): e39632.
Schwarz D S, Hutvagner G, Du T, Xu Z, Aronin N and Zamore P D (2003). "Asymmetry in the assembly of the RNAi enzyme complex." Cell 115(2): 199-208.
Seo G J, Chen C J and Sullivan C S (2009). "Merkel cell polyomavirus encodes a microRNA with the ability to autoregulate viral gene expression." Virology 383(2): 183-187.
Seo G J, Fink L H, O'Hara B, Atwood W J and Sullivan C S (2008). "Evolutionarily conserved function of a viral microRNA." J Virol 82(20): 9823-9828.
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(11): 881-886.
Shai A, Brake T, Somoza C and Lambert P F (2007). "The human papillomavirus E6 oncogene dysregulates the cell cycle and contributes to cervical carcinogenesis through two independent activities." Cancer Res 67(4): 1626-1635.
Shen S N, Wang L F, Jia Y F, Hao Y Q, Zhang L and Wang H (2013). "Upregulation of microRNA-224 is associated with aggressive progression and poor prognosis in human cervical cancer." Diagn Pathol 8: 69.
Shenouda S K and Alahari S K (2009). "MicroRNA function in cancer: oncogene or a tumor suppressor?" Cancer Metastasis Rev 28(3-4): 369-378.
Skalsky R L and Cullen B R (2010). "Viruses, microRNAs, and host interactions." Annu Rev Microbiol 64: 123-141.
Smyth G K (2004). "Linear models and empirical bayes methods for assessing differential expression in microarray experiments." Stat Appl Genet Mol Biol 3: Article3.
Sonkoly E, Ståhle M and Pivarcsi A (2008). "MicroRNAs: novel regulators in skin inflammation." Clin Exp Dermatol 33(3): 312-315.
Sonkoly E, Wei T, Janson P C, Sääf A, Lundeberg L, Tengvall-Linder M, Norstedt G, Alenius H, Homey B*, et al.* (2007). "MicroRNAs: novel regulators involved in the pathogenesis of psoriasis?" PLoS One 2(7): e610.
Stark M S, Tyagi S, Nancarrow D J, Boyle G M, Cook A L, Whiteman D C, Parsons P G, Schmidt C, Sturm R A*, et al.* (2010). "Characterization of the Melanoma miRNAome by Deep Sequencing." PLoS One 5(3): e9685.
Stöppler M C, Straight S W, Tsao G, Schlegel R and McCance D J (1996). "The E5 gene of HPV-16 enhances keratinocyte immortalization by full-length DNA." Virology 223(1): 251-254.
Straight S W, Hinkle P M, Jewers R J and McCance D J (1993). "The E5 oncoprotein of human papillomavirus type 16 transforms fibroblasts and effects the downregulation of the epidermal growth factor receptor in keratinocytes." J Virol 67(8): 4521-4532.
Sullivan C S, Grundhoff A T, Tevethia S, Pipas J M and Ganem D (2005). "SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells." Nature 435(7042): 682-686.
Taganov K D, Boldin M P, Chang K J and Baltimore D (2006). "NF-kappaB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses." Proc Natl Acad Sci U S A 103(33): 12481-12486.
Tang T, Wong H K, Gu W, Yu M Y, To K F, Wang C C, Wong Y F, Cheung T H, Chung T K*, et al.* (2013). "MicroRNA-182 plays an onco-miRNA role in cervical cancer." Gynecol Oncol 129(1): 199-208.

Thomson J M, Parker J, Perou C M and Hammond S M (2004). "A custom microarray platform for analysis of microRNA gene expression." Nat Methods 1(1): 47-53.

Tomakidi P, Cheng H, Kohl A, Komposch G and Alonso A (2000). "Modulation of the epidermal growth factor receptor by the human papillomavirus type 16 E5 protein in raft cultures of human keratinocytes." Eur J Cell Biol 79(6): 407-412.

Valencia-Sanchez M A, Liu J, Hannon G J and Parker R (2006). "Control of translation and mRNA degradation by miRNAs and siRNAs." Genes Dev 20(5): 515-524.

Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek J A, Costa G*, et al.* (2008). "A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning." Genome Res 18(7): 1051-1063.

van der Burgt A, Fiers M W, Nap J P and van Ham R C (2009). "In silico miRNA prediction in metazoan genomes: balancing between sensitivity and specificity." BMC Genomics 10: 204.

Veronese A, Lupini L, Consiglio J, Visone R, Ferracin M, Fornari F, Zanesi N, Alder H, D'Elia G*, et al.* (2010). "Oncogenic role of miR-483-3p at the IGF2/483 locus." Cancer Res 70(8): 3140-3149.

Visone R, Rassenti L Z, Veronese A, Taccioli C, Costinean S, Aguda B D, Volinia S, Ferracin M, Palatini J*, et al.* (2009). "Karyotype-specific microRNA signature in chronic lymphocytic leukemia." Blood 114(18): 3872-3879.

Visone R, Russo L, Pallante P, De Martino I, Ferraro A, Leone V, Borbone E, Petrocca F, Alder H*, et al.* (2007). "MicroRNAs (miR)-221 and miR-222, both overexpressed in human thyroid papillary carcinomas, regulate p27Kip1 protein levels and cell cycle." Endocr Relat Cancer 14(3): 791-798.

Visone R, Veronese A, Rassenti L Z, Balatti V, Pearl D K, Acunzo M, Volinia S, Taccioli C, Kipps T J*, et al.* (2011). "miR-181b is a biomarker of disease progression in chronic lymphocytic leukemia." Blood 118(11): 3072-3079.

Volinia S, Calin G A, Liu C G, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C*, et al.* (2006). "A microRNA expression signature of human solid tumors defines cancer gene targets." Proc Natl Acad Sci U S A 103(7): 2257-2261.

Volinia S, Galasso M, Costinean S, Tagliavini L, Gamberoni G, Drusco A, Marchesini J, Mascellani N, Sana M E*, et al.* (2010). "Reprogramming of miRNA networks in cancer and leukemia." Genome Res 20(5): 589-599.

Volinia S, Visone R, Galasso M, Rossi E and Croce C M (2010). "Identification of microRNA activity by Targets' Reverse EXpression." Bioinformatics 26(1): 91-97.

Wang W C, Lin F M, Chang W C, Lin K Y, Huang H D and Lin N S (2009). "miRExpress: analyzing high-throughput sequencing data for profiling microRNA expression." BMC Bioinformatics 10: 328.
Wang X, Meyers C, Wang H K, Chow L T and Zheng Z M (2011). "Construction of a full transcription map of human papillomavirus type 18 during productive viral infection." J Virol 85(16): 8080-8092.
Wang X, Tang S, Le S Y, Lu R, Rader J S, Meyers C and Zheng Z M (2008). "Aberrant expression of oncogenic and tumor-suppressive microRNAs in cervical cancer is required for cancer cell growth." PLoS One 3(7): e2557.
Wang X, Wang H K, McCoy J P, Banerjee N S, Rader J S, Broker T R, Meyers C, Chow L T and Zheng Z M (2009). "Oncogenic HPV infection interrupts the expression of tumor-suppressive miR-34a through viral oncoprotein E6." RNA 15(4): 637-647.
Wang Z, Gerstein M and Snyder M (2009). "RNA-Seq: a revolutionary tool for transcriptomics." Nat Rev Genet 10(1): 57-63.
Werness B A, Levine A J and Howley P M (1990). "Association of human papillomavirus types 16 and 18 E6 proteins with p53." Science 248(4951): 76-79.
Wheeler D A, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen Y J, Makhijani V*, et al.* (2008). "The complete genome of an individual by massively parallel DNA sequencing." Nature 452(7189): 872-876.
Williamson V, Kim A, Xie B, McMichael G O, Gao Y and Vladimirov V (2013). "Detecting miRNAs in deep-sequencing data: a software performance comparison and evaluation." Brief Bioinform 14(1): 36-45.
Wilting S M, Snijders P J, Verlaat W, Jaspers A, van de Wiel M A, van Wieringen W N, Meijer G A, Kenter G G, Yi Y*, et al.* (2013). "Altered microRNA expression associated with chromosomal changes contributes to cervical carcinogenesis." Oncogene 32(1): 106-116.
Wilting S M, van Boerdonk R A, Henken F E, Meijer C J, Diosdado B, Meijer G A, le Sage C, Agami R, Snijders P J*, et al.* (2010). "Methylation-mediated silencing and tumour suppressive function of hsa-miR-124 in cervical cancer." Mol Cancer 9: 167.
Witten D, Tibshirani R, Gu S G, Fire A and Lui W O (2010). "Ultra-high throughput sequencing-based small RNA discovery and discrete statistical biomarker analysis in a collection of cervical tumours and matched controls." BMC Biol 8: 58.
Woodman C B, Collins S I and Young L S (2007). "The natural history of cervical HPV infection: unresolved issues." Nat Rev Cancer 7(1): 11-22.
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(9): 1710-1716.

Xie H, Zhao Y, Caramuta S, Larsson C and Lui W O (2012). "miR-205 expression promotes cell proliferation and migration of human cervical cancer cells." PLoS One 7(10): e46990.

Xu N, Segerman B, Zhou X and Akusjarvi G (2007). "Adenovirus virus-associated RNAII-derived small RNAs are efficiently incorporated into the rna-induced silencing complex and associate with polyribosomes." J Virol 81(19): 10540-10549.

Yamamoto N, Kinoshita T, Nohata N, Itesako T, Yoshino H, Enokida H, Nakagawa M, Shozu M and Seki N (2013). "Tumor suppressive microRNA-218 inhibits cancer cell migration and invasion by targeting focal adhesion pathways in cervical squamous cell carcinoma." Int J Oncol 42(5): 1523-1532.

Yang J S, Phillips M D, Betel D, Mu P, Ventura A, Siepel A C, Chen K C and Lai E C (2011). "Widespread regulatory activity of vertebrate microRNA* species." RNA 17(2): 312-326.

Yang Z, Chen S, Luan X, Li Y, Liu M, Li X, Liu T and Tang H (2009). "MicroRNA-214 is aberrantly expressed in cervical cancers and inhibits the growth of HeLa cells." IUBMB Life 61(11): 1075-1082.

Yao Q, Xu H, Zhang Q Q, Zhou H and Qu L H (2009). "MicroRNA-21 promotes cell proliferation and down-regulates the expression of programmed cell death 4 (PDCD4) in HeLa cervical carcinoma cells." Biochem Biophys Res Commun 388(3): 539-542.

Yu J, Wang Y, Dong R, Huang X, Ding S and Qiu H (2012). "Circulating microRNA-218 was reduced in cervical cancer and correlated with tumor invasion." J Cancer Res Clin Oncol 138(4): 671-674.

Zhang B, Pan X, Cobb G P and Anderson T A (2007). "microRNAs as oncogenes and tumor suppressors." Dev Biol 302(1): 1-12.

Zhang W, Gao S, Zhou X, Xia J, Chellappan P, Zhou X, Zhang X and Jin H (2010). "Multiple distinct small RNAs originate from the same microRNA precursors." Genome Biol 11(8): R81.

Zheng Z M and Baker C C (2006). "Papillomavirus genome structure, expression, and post-transcriptional regulation." Front Biosci 11: 2286-2302.

Zhou X, Chen X, Hu L, Han S, Qiang F, Wu Y, Pan L, Shen H, Li Y*, et al.* (2010). "Polymorphisms involved in the miR-218-LAMB3 pathway and susceptibility of cervical cancer, a case-control study in Chinese women." Gynecol Oncol 117(2): 287-290.

Zhu E, Zhao F, Xu G, Hou H, Zhou L, Li X, Sun Z and Wu J (2010). "mirTools: microRNA profiling and discovery based on high-throughput sequencing." Nucleic Acids Res 38(Web Server issue): W392-397.

zur Hausen H (2000). "Papillomaviruses causing cancer: evasion from host-cell control in early events in carcinogenesis." J Natl Cancer Inst 92(9): 690-698.

zur Hausen H (2002). "Papillomaviruses and cancer: from basic studies to

clinical application." Nat Rev Cancer 2(5): 342-350.
zur Hausen H (2009). "Papillomaviruses in the causation of human cancers - a brief historical account." Virology 384(2): 260-265.