miRNAs in equine sarcoids: Identification and profiling of miRNAs in equine fibroblasts and BPV-1 transformed equine fibroblasts

Dr Nuria Terrón Canedo

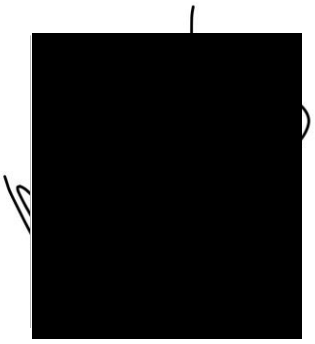
September 2016

This thesis is submitted to the University of Glasgow in accordance with the requirements for the degree of Doctor in Philosophy in the School of Veterinary Medicine.

Author's declaration:

I declare that the work carried out in this thesis is original and it was carried out by either myself or with the appropriate acknowledgements. Exceptions to this have been the miRCURY LNA™ microRNA Array 7th Gen (performed by an external company, Exiqon A/S, Vedbeck, Denmark), high throughput sequencing (performed by Miss Julie Galbraith at Glasgow Polyomics, University of Glasgow) and several parts of the bioinformatic analysis such as set-up of the miRDeep 2.0. pipeline and creation of scripts for data analysis of high throughput sequencing (performed by Dr. Willie Weir, University of Glasgow).

The work contained in this thesis has not been presented for the award of a degree in any other university and it was carried out in the University of Glasgow under the co-supervision of Prof. L. Nasir and Dr. C. Britton.



Nuria Terron Canedo

September 2016

Acknowledgments

I would like to thank my supervisors Prof Lubna Nasir and Dr Collette Britton for their help and guidance in the different parts of this project. I would also like to thank them for giving me both support in the lab and outside of the lab, when sometimes it was most needed. I would like to express my eternal gratitude to The Horserace Betting Levy Board (HBLB) and The James Herriot Scholarship Fund for the financial support, without which this project would have never happened. I would also like to thank Dr Lesley Nicolson and Dr William Weir for their contribution to the project. Special thanks to Dr William Weir for his invaluable advice on the thesis as well as on my future career. I would like to thank Miss Lizzie Gault, Dr Alan Winter, Dr Idoia Busnadiego, Mr Andrew Stevenson, Miss Natalie Hutchinson and of course my office mates Dr Adam Bell and Dr Elspeth Waugh for sharing their experience in numerous laboratory techniques that were completely new to me at the start of my PhD and became my 'bread-and-butter'. Thanks to Dr Marco Duz for his friendship and support as a PhD sufferer and for all the de-stress therapy sessions in the form of dog walks.

Huge thanks to my family, who have been incredibly supportive throughout my career and studies. Particularly I'd like to thank my mum, who has always believed in me, has encouraged me to be a curious mind and has taught me the lesson that knowledge and education are the best gifts you can ever give to a daughter. Thanks to my granny, who I believe I get the stubbornness and perseverance from, that allowed me to put the effort and hard-work during this PhD. (Spanish translation: Quisiera expresar mi enorme gratitud a mi familia, que siempre me han dado su apoyo incondicional a lo largo de las etapas tan diversas de mi carrera y estudios. En particular, me gustaria darle las gracias a mi madre, que siempre ha creido en mi, ha estimulado my curiosidad me ha enseñado que el conocimiento y la educación son los mejores regalos que se le pueden dar a una hija. Gracias a mi abuela, de quien creo que he heredado la cabezonería y perserverancia, que me han permitido poner el esfuerzo y horas de trabajo durante este doctorado). Last, but not least, big thanks to my all of my friends here in Glasgow and to all of those far away from Glasgow, who all have contributed to help me to become who I am, and who have made the journey of my PhD one of the best experiences in my life.

Abstract

Introduction: Equine sarcoids (ES) are fibroblastic cutaneous tumours, affecting members of the *Equidae* family (horses, donkeys, zebras), hybrids (mules) and other ungulates (tapirs and giraffes) (Marti *et al.*, 1993; Reid *et al.*, 1994; Knottenbelt *et al.*, 1995; Nel *et al.*, 2006; Valentine, 2006; Kidney & Berrocal, 2008; van Dyk *et al.*, 2009; van Dyk *et al.*, 2011; Schaffer *et al.*, 2013). This debilitating disease has a prevalence of 5.6% in the British equestrian population and it is a major welfare concern in less developed countries (Ayele *et al.*, 2007; NEHS, 2015). Age (younger horses), genetic predisposition (presence of equine leucocyte antigens class I A3 and class II B1; regions of the equine genome in chromosomes 20, 22, 23, 25 with microsatellite markers, single nucleotide polymorphisms or deletions) and breed (Quarter horse and Thoroughbred) have been identified as risk factors (Lazary *et al.*, 1985; Brostrom *et al.*, 1988; Mohammed *et al.*, 1992; Marti *et al.*, 1993; Wobeser *et al.*, 2010; Jandova *et al.*, 2012; Bugno-Poniewierska *et al.*, 2016; Staiger *et al.*, 2016). Diagnosis and treatment are challenging due to the variable clinical presentation (six clinical types), inconsistent histopathological findings and gaps in the understanding of the aetiology (Martens *et al.*, 2000; Knottenbelt, 2005b; Taylor & Haldorson, 2013). Bovine papillomavirus type 1 and 2 (BPV-1/-2) are intrinsically linked to the development of ES. Episomal viral genomes (from 2 to over 100 copies per cell) as well as expression of the viral oncoprotein E5, E6 and E7 have been found in clinical lesions in several studies (Amtmann *et al.*, 1980; Angelos *et al.*, 1991; Nasir & Reid, 1999; Carr *et al.*, 2001b; Chambers *et al.*, 2003b; Ambros, 2004; Bogaert *et al.*, 2007; Yuan *et al.*, 2007b; Borzachiello *et al.*, 2008; Nasir & Campo, 2008; Haralambus *et al.*, 2010; Brandt *et al.*, 2011b; Nasir & Brandt, 2013; Wilson *et al.*, 2013). Activation of the platelet-derived growth factor β receptor (PDGFβ-R) signaling pathway and the intracellular retention of major histocompatibility complex (MHC) class I, both driven by the BPV-1/-2 E5 oncoprotein, are the most studied cellular events in tumoural transformation in equine sarcoids (Petti *et al.*, 1991; Ashrafi *et al.*, 2002; Borzacchiello *et al.*, 2006; Borzachiello *et al.*, 2008; Marchetti *et al.*, 2009; Altamura *et al.*, 2013). Other recognised effects of BPV-1/-2 oncoproteins in ES are activation of p38 mitogen-activated protein kinases (p38-MAPK) (Yuan *et al.*, 2011c), intracellular mislocation of tumour suppressor p53 (Yuan *et al.*, 2008a), activation of p600, a member of the retinoblastoma family (Corteggio *et al.*, 2011), downregulation of

toll-like receptor 4 (TLR4) (Yuan *et al.*, 2010a) and increased expression of forkhead box protein 3 (FOXP3). However, to date, the mechanisms by which the virus activates or inhibits these pathways are still poorly understood. It has been shown that in human papillomavirus (HPV) induced cancer, multiple oncogenic pathways are regulated by microRNAs (miRNAs), which are small non-coding RNAs and essential regulators at a cellular level (Gómez-Gómez *et al.*, 2013; Honegger *et al.*, 2015). The aberrant expression of miRNAs found in high-risk HPVs-infected cells (HR-HPV) and cervical cancer samples has contributed to the understanding of the mechanisms by which the virus initiates and perpetuates malignant transformation. Dysregulation of miRNAs (miR-106b, miR-145, miR-15/16 cluster, miR-196a, miR-21, miR-23b and miR-92 amongst other miRNAs) has been attributed to the presence of the HPV oncoproteins E6 and E7 (Lajer *et al.*, 2012; Ben *et al.*, 2015; Honegger *et al.*, 2015). Currently in the official database of miRNAs, miRBase 20.0 (www.mirbase.org) there are 711 annotated equine miRNA precursors, encoding 690 mature miRNAs, of which more than half are predicted sequences (359) (Zhou *et al.*, 2009) and the rest have been verified in one equine testicular sample with the use of high throughput sequencing (HTS) (331 sequences) (Platt *et al.*, 2014). Nonetheless, other studies attempted to describe miRNAs in the horse using HTS in equine semen, bone, cartilage, colon, liver and muscle (Das *et al.*, 2013; Desjardin *et al.*, 2014; Kim *et al.*, 2014) although these are not included in the miRBase 20.0.

Aims: The goals of the present study were to produce a detailed catalogue of mature sequences expressed in equine normal fibroblast cells (equine primary fibroblasts) and study the differential expression of miRNAs in equine sarcoids using an *in vitro* model previously developed at the University of Glasgow (Yuan *et al.*, 2008a).

Materials and methods: RNA was extracted from three cell lines from an equine *in vitro* sarcoid model: normal equine fibroblasts (EqPalFs); cells transformed with the BPV-1 genome, EqPalF-BPV-1 cells referred to as S6-2 cells; and equine tumoural fibroblasts, named EqSO4bs (Yuan et al., 2008b). miRNAs contained in the RNA fraction were assessed using miRNA microarray (EqPalF *vs* S6-2 cells), high throughput sequencing (HTS) (EqPalF, EqSO4b and S6-2 cells) and real-time reverse transcription polymerase chain reaction (qRT-PCR). An adjusted p-value

of less than 0.05 was taken to be statistically significant when studying the differential expression of miRNAs in the three cell lines.

Results: Out of the 2164 probes in the microarray, once quality probes were removed (90 probes), 402 probes had signal above the background level, representing 402 mature miRNA sequences present in equine fibroblasts. Additionally, the results confirmed that 124 of the equine predicted miRNAs currently in the official miRBase are indeed encoded within the equine genome and expressed in equine fibroblasts. In total 206 miRNAs were found to be significantly dysregulated in BPV-transformed equine primary fibroblasts (S6-2 cells) when compared to control primary fibroblasts (EqPalF) using miRNA microarray (adjusted $p<0.05$). Dysregulated miRNAs of interest in BPV-1 transformed cells captured in the microarray results were let-7b, miR-106a, miR-23b, miR-143 and miR-181a.

With the use of HTS, 593 putative miRNA candidates in EqPalF cells were described providing experimental verification of the existence of 258 equine miRNAs and the presence of an additional 335 novel miRNAs in the horse (162 were homologous to miRNAs identified in other species and 173 have not been reported before in the horse or other species). Descriptive parameters (sequence length, nucleotide composition, number of miRNAs in each chromosome, miRNA abundance, ratio 3p/5p) were found similar to previous studies except a higher number of miRNAs found in equine chromosome 24 (ECA24). Amongst the expressed miRNAs in equine primary fibroblasts, the miRNAs provisionally named as eca-miR-21, eca-miR-22, eca-miR-26a, eca-miR-27b, eca-miR-28-3p, eca-miR-92a, eca-miR-127, eca-miR-143, eca-miR-181a and eca-miR-191a were the 10 most abundantly in equine primary fibroblasts. Moreover, with the use of HTS and bioinformatic tools miRNAs additional 22 equine predicted miRNAs and additional 112 novel miRNAs were identified in the BPV-1 positive cells that were not expressed in the equine primary fibroblasts. When studying the differences between cell lines, 201 miRNAs were linked to BPV-1 presence, 128 to tumoural transformation and 106 had the potential of biomarkers for the presence of BPV-1 and/or tumoural development. Interestingly the degree of correlation of miRNA read counts between BPV-1 positive cells (EqSO4b and S6-2) was very high (Pearson's coefficient of 0.91). Novel miRNAs will be submitted to miRBase (http://mirbase.org/) to confirm

their identity and they will represent an important contribution to the scientific community as a resource for further studies on equine miRNAs. Amongst the dysregulated miRNAs, miR-181a and miR-181b were found to be downregulated by both HTS and microarray in BPV-1 positive cells and miR-181a showed the same pattern by qRT-PCR. Importantly, low levels of miR-181a were flagged up by the biomarker algorithm as a suitable biomarker for BPV-1 presence and were also encoded in areas of the equine genome previously reported to be linked to ES (Jandova *et al.*, 2012; Bugno-Poniewierska *et al.*, 2016). Interestingly, the human homologues are linked to HR-HPV carcinogenesis (Lee *et al.*, 2015). Other miRNAs of interest were miR-196a/b that were found upregulated only in tumoural cells and were reported to be associated with increased levels of MMPs through activation of NME4-JNK-TIMP1-MMPs pathway in oral cancer, and in colon cancer (Lu *et al.*, 2014; Mo *et al.*, 2015). Other dysregulated miRNAs that were also predicted biomarkers for BPV-1 presence were miR-92a (member of the miR-17~92 cluster), miRNAs from the 15/16 cluster and miR-21. Dysregulated miRNAs in BPV-1 positive cells were found to be in networks related to cellular growth, cancer, and HPV related networks.

Conclusions: With the use of several molecular biology and bioinformatic tools, 280 previously predicted equine miRNAs and 285 novel miRNAs were demonstrated in equine fibroblasts. Presence of BPV-1 is linked to 201 dysregulated miRNAs and 128 miRNAs were intrinsically linked to tumoural transformation in ES. These results will be an addition to the scientific community when deposited in miRBase and have contributed to a better understanding of miRNAs in the horse and BPV-1 oncogenesis in ES.

Table of Contents

List of Figures

List of Tables

Publications

Terron-Canedo, N., Britton, C., Nicolson, L., & Nasir, L. (2014). Differential Expression of microRNAs in Equine Sarcoids: Preliminary Results from an *In Vitro* Model. *Equine Vet J*, 46: 15. doi:10.1111/evj.12323_32.

Terron-Canedo, N., Weir, W., Nicolson, L., Britton, C., & Nasir, L. (2016). Differential expression of microRNAs in bovine papillomavirus type 1 transformed equine cells. *Vet Comp Oncol*. doi:10.1111/vco.12216.

Abbreviations / definitions

°C	Degrees centigrade
µg	Microgram
µl	Microlitre
AKT	Protein kinase B
BCG	Bacille Calmette Guerin, an immunotherapy used for the treatment of ES
bp	Base pairs
BPVs	Bovine papillomaviruses. Example: BPV-1, BPV-2
CBP	Crebb binding protein
c-src	Proto-oncogene tyrosine-protein kinase
DGCR8	Protein DiGeorge syndrome chromosomal (or critical) region 8
DMEM	Dulbecco's modified eagle's medium
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTPs	Deoxyribonucleic thriphosphates
dsDNA	Double stranded DNA
DQA1	Gene DQ alpha 1 (MHC class II gene in the ECA 20)
ECA	equine chromosome
EDEM2	endoplasmic reticulum degradation-enhancing a-mannosidase I-like receptor (ECA 22)
EGF	epidermal growth factor
EIF6	eukaryotic translation initiation factor 6 (ECA 22),
Erk	Extracellular- signal – regulated kinase
ECA	Equine chromosome
ELA	Equine leucocyte antigen, corresponding to the genes that encode for the major histocompatibility complex in the horse
ES	Equine sarcoids
GRB2	Growth-factor-receptor-bound protein 2
HPVs	Human papillomaviruses
HR-HPVs	High-risk human papillomaviruses, predominantly HPV-16/-18
HTS	High throughput sequencing
Kbp	Kilobase pairs

L	Litre
LCR	Long control region of genome in PVs
M	Molar
MAML1	Mastermind-like protein 1 when it binds BPV-1 E6 it represses the NOTCH pathway and consequent gene transcription
Mek or MAPK	Microtubule associated protein kinase, mitogen-activated protein kinase p38
mg	Milligram
MHC	Major histocompatibility complex
ml	Millilitre
mM	Millimolar
MMPs	Matrix metalloproteinases. Example: MMP-1, MMP-24
mRNA	Messenger RNA
NOS	Naturally occurring equine sarcoids
NEHS	National Equine Health Survey
ORF	Open reading frame
p600	600-kDa retinoblastoma protein associated factor
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDGFß-R	Platelet-derived growth factor receptor
PI3K	Phosphatidylinositol – 3 – kinase
pmol	Picomol
pRB	Retinoblastoma protein
PROCR	Endothelial protein C receptor (ECA 22)
PVs	Papillomaviruses
qRT-PCR	(Real-time) reverse-transcription polymerase chain reaction
QTL	quantitative trait locus
Ras-MAPK	Ras-mitogen activated protein kinase
RNA	Ribonucleic acid
RISC	RNA-induced silencing complex
SNPs	single nucleotide polymorphisms
SOS	‘son of sevenless’ gene protein
TERT	telomerase reverse transcriptase
TNF	Tumour necrosis factor
V-ATPase	Vacuolar proton-ATPase 16-kDa subunit

Chapter 1. Introduction to equine sarcoids and miRNAs in the horse

1.1. Introduction to equine sarcoids

1.1.1. Definition of equine sarcoids

Equine sarcoids (ES) are locally aggressive fibroblastic cutaneous tumours affecting members of the *Equidae* family (horses, donkeys, zebras and mules) (Marti *et al.*, 1993; Reid *et al.*, 1994; Knottenbelt *et al.*, 1995; Nel *et al.*, 2006; Valentine, 2006; van Dyk *et al.*, 2009; Schaffer *et al.*, 2013). Sarcoids also have been described in non-equids such as tapirs and giraffes (Kidney & Berrocal, 2008; van Dyk *et al.*, 2011; Schaffer *et al.*, 2013). Equine sarcoids were firstly reported in 1936 as fibro-epithelial tumours affecting horses in South Africa and the pattern of spread suggested an infectious origin (Jackson, 1936). They are classified by the World Health Organisation as tumours of fibrous tissue within the group of mesenchymal-soft tissue tumours (Weiss, 1974).

The term 'sarcoid', which in Greek means 'a disease that resembles crude flesh', was transferred to veterinary medicine from human sarcoidosis (Boeck's disease). However, distinction should be made to human 'sarcoidosis', a rare condition of unknown aetiology affecting the immune system with formation of visceral granulomas derived from the reticuloendothelial system, causing respiratory and other clinical signs that differ from equine sarcoids (Hunninghake *et al.*, 1999).

1.1.2. Epidemiology of equine sarcoids

1.1.2.1. Prevalence

Equine sarcoids are the most common cutaneous tumours in horses (Schaffer *et al.*, 2013). Prevalence rates range between 0.4-15.8% of all clinical referral cases (Lory *et al.*, 1993; Valentine, 2006) and from 11.7% to 67% of all equine neoplasms (Marti *et al.*, 1993; Valentine, 2006; Schaffer *et al.*, 2013; Taylor & Haldorson, 2013). In the United Kingdom and USA, equine sarcoids are amongst the most common equine cutaneous diseases (Schaffer *et al.*, 2013; NEHS, 2014). The most recent epidemiological study in the United Kingdom, based on the questionnaire responses from 4951 owners in reference to 14952 horses, shows a

prevalence of equine sarcoids of 5.6% in the British equine population (NEHS, 2015). The five most common skin diseases reported by horse owners during this study are presented in Figure 1.1.

In wild zebras, prevalence of ES is higher (53%) although this may be a reflection of an elevated degree of inbreeding in this species (Zimmerman, 2004; Nel *et al.*, 2006; Marais *et al.*, 2007).

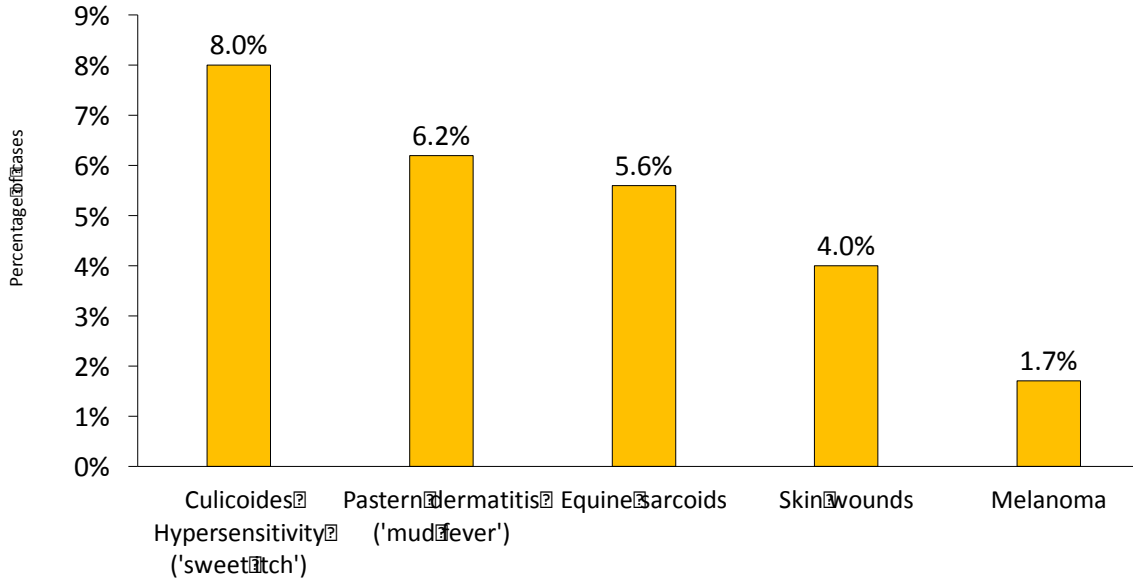


Figure 1.1. Prevalence of the five most common equine skin diseases reported in the British equine population. Data from 14952 horses, (NEHS, 2015)

Equine sarcoids are seldom life threatening but they have an effect on equine health and welfare as well as the material value of a horse (Gerber, 1989; Marti *et al.*, 1993). The presence of sarcoid tumours decreases the commercial value of an equid during pre-purchase examination and it can lead to a horse being excluded from an insurance policy. Bacterial infection, difficulties in defecation, urination and locomotion and mastication have been reported to be secondary consequences of sarcoids in working horses (Ayele *et al.*, 2007).

1.1.2.2. Risk factors in equine sarcoids

Gender and age

Although the earliest studies found that younger males were at higher risk, there is no consistent evidence to support gender predilection in equine sarcoids (Mohammed *et al.*, 1992; Reid *et al.*, 1994; Torrontegui & Reid, 1994; Wobeser *et al.*, 2010). Sarcoids have been reported to develop in skin discontinuities and

areas of the body subjected to trauma (Voss, 1969; Torrontegui & Reid, 1994) and flies have been attributed a role in the dissemination of the disease (Finlay *et al.*, 2009). These findings, combined with the fact that male neutering with healing by second intention is a common practice (Kilcoyne *et al.*, 2013), may facilitate the entrance in the peri-inguinal area in recently castrated young horses of the bovine papillomavirus type 1/-2, virus associated with the development of equine sarcoids (for more details on aetiology please refer to section 1.1.4.1, current Chapter). This may have led to the belief that this group are at higher risk. A recent study enlisting 302 equids has shown no statistical evidence that geldings are at higher risk of developing ES compared to stallions or mares (Wobeser *et al.*, 2010).

Studies found that prevalence of sarcoids is indeed higher in younger animals and decreases with age after a peak at 15 years old (Mohammed *et al.*, 1992; Marti *et al.*, 1993; Reid & Mohammed, 1997; Valentine, 2006). Equine sarcoids can develop in equids as young as 6 months old (Ayele *et al.*, 2007; Wobeser *et al.*, 2010). The mean age, despite variability across studies, is around 3.5-4 years old (Miller & Campbell, 1982; Marti *et al.*, 1993; Torrontegui & Reid, 1994; Brostrom, 1995; Scott *et al.*, 2001). The latest epidemiological study carried out in Canada over 12 years in 686 horses shown that those biopsied due to the presence of a sarcoid were significantly younger than those biopsied due to other skin conditions (Wobeser *et al.*, 2010). To explain this age distribution, it has been hypothesised that older horses may develop immunity to ES and that genetically predisposed animals suffer from neoplasias at an earlier age (Torrontegui & Reid, 1994).

Genetic predisposition

The major histocompatibility complex (MHC) has been reported as a risk factor in equine sarcoids. The MHC contains genes that encode for lymphocyte-presenting glycoproteins and other molecules that participate in the immune adaptive and innate responses. MHC encoded molecules are conventionally classified as class I, II and III. In *Equus caballus*, the genes that encode for the equine MHC, named equine leucocyte antigens (ELA) are located on chromosome 20 (Ansari *et al.*, 1988).

ELA haplotypes A3 (class I, previously named as W3) and B1 (class II, previously named as W13) have been linked to susceptibility to sarcoids in tests performed on 134 horses from different breeds (mainly Swiss Warmblood, Irish hunter, French Warmblood and Freiberger) (Lazary *et al.*, 1985). This has been confirmed for Swedish half-bred horses (Brostrom *et al.*, 1988; Brostrom, 1995) and Thoroughbred horses (Meredith *et al.*, 1986), where the autosomal ELA-linked genes for class I A3 and class II B1 were strongly associated with sarcoid development. ELA class II B1 (W13) has not been detected in Standardbreds, one of the breeds with the lowest incidence of ES (Meredith *et al.*, 1986). The link between the presence of certain ELA antigens and the risk of developing equine sarcoids is still not clear, as not all horses with haplotype B1 suffer from sarcoids (Goodrich *et al.*, 1998).

Quantitative trait locus (QTL) mapping has identified regions within chromosomes 20, 23 and 25 linked with increased susceptibility to equine sarcoids in a population of 26 Swiss Warmblood horses (Jandova *et al.*, 2012). Interestingly, the microsatellite markers linked to increased risk for ES on equine chromosome 20 were mapped outside of the ELA region. More recently, allelic association tests, using genome wide microchip single nucleotide polymorphisms (SNPs), in a study conducted on 59 horses with clinically diagnosed sarcoids, revealed that the phenotype associated with presence of ES may be linked to polymorphisms found on chromosome 20 (20q28-20q33, within the region of ELA genes) and chromosome 22 (22q24-22q26). Highly significant SNPs mapped to these two genomic regions encoded the gene DQ alpha 1 (DQA1) an MHC class II gene (chromosome 20), matrix metalloproteinase 24 (MMP-24) (chromosome 22), eukaryotic translation initiation factor 6 (EIF6) (chromosome 22), endoplasmic reticulum degradation-enhancing a-mannosidase I-like receptor (EDEM2) (chromosome 22) and endothelial protein C receptor (PROCR) (chromosome 22) (Staiger *et al.*, 2016).

Comparative genomic hybridisation was used recently to assess regions of deletion or amplification linked to ES. In tumoural cells from naturally occurring sarcoids (NOS) there is a significant deletion in equine chromosome 23 (23q18-23q19) and duplications of several telomere and rDNA regions (Bugno-Poniewierska *et al.*, 2016).

It is unclear whether ELA types and loci identified as being linked to sarcoid development are virus permissive or whether they lead to an impaired immune response that facilitates the development of sarcoids (please refer to section 1.1.4.1 for more details, current Chapter). It is likely that susceptibility to ES is driven by multiple genes (Christen *et al.*, 2014).

Breed

It is well established that Standardbred horses are less likely to suffer from sarcoids compared to other breeds (Meredith *et al.*, 1986; Angelos *et al.*, 1988; Mohammed *et al.*, 1992), although a recent study placed Warmblood horses at a lower risk when compared to Standardbred horses (Wobeser *et al.*, 2010). The disparity between Wobeser *et al.* (2010) and previous studies could be due to early studies not accounting for Warmblood horses as a separate group. Quarter horses on the other hand are consistently quoted within the top three breeds with higher risk and incidence of ES (Angelos *et al.*, 1991; Mohammed *et al.*, 1992; Wobeser *et al.*, 2010). Arabs, Appaloosas and Thoroughbreds have been reported to be within the top five breeds with higher risk of developing sarcoids, and the placement depends on the study and the breeds included in the study (Angelos *et al.*, 1991; Mohammed *et al.*, 1992; Wobeser *et al.*, 2010). Donkeys on the other hand have a much higher risk of suffering ES compared to horses (Wobeser *et al.*, 2010).

Anatomical region

Equine sarcoids can affect any part of the body. Typically, they develop in the head (periorbital and ear), neckline, axilla, distal limbs, ventral abdomen (paragenital region) and groin (Ragland *et al.*, 1970a; Goodrich *et al.*, 1998; Martens *et al.*, 2000). In Australia, USA and mainland Europe isolated or a small number of tumours have been reported, whereas in the U.K. affected animals more often than not have multiple lesions (Knottenbelt *et al.*, 1995). Previous history of trauma or skin damage has been suggested to be a factor that increases susceptibility to ES (Ragland *et al.*, 1970a; Knottenbelt, 2005b).

1.1.3. Clinical aspects of equine sarcoids

1.1.3.1. Equine sarcoid subtypes

Equine sarcoids have a variable clinical presentation. They can present as unique masses or multiple lesions. They are not painful and also non-pruritic unless ulceration and secondary infection occur (Hamann & Grabner, 2005). Classification of sarcoids is based on macroscopic appearance and skin involvement (Knottenbelt, 2005b). This classification is useful for clinical decisions, for a treatment plan, to monitor progress and prognosis. In brief, equine sarcoids are classified as occult, verrucose, nodular (type A1, type A2, type B1 and type B2), fibroblastic (type 1a, type 1b and type 2), mixed and malignant. Examples are displayed in Figure 1.2. Occult sarcoids are the most benign form of ES and are thought to represent an early stage of the disease. Occult sarcoids are often circular alopecic lesions with a grey colour, that lack hair and their appearance is of a hyperkeratotic dry lesion. On palpation, skin feels thickened, dry and has small nodules underneath. This form is easily misdiagnosed as 'ring-worm' (dermatophytosis) or skin rubs from tack or other equipment (see Figure 1.2, Image 1). Verrucose sarcoids are prominent from the skin and they are similar to warts in appearance. The skin on these lesions is dry, alopecic and scaly. The margins are not well defined. Due to the 'wart-like' appearance these can be mistaken for equine papillomas. Nodular sarcoids are circumscribed firm subcutaneous masses that can appear as single or multiple lesions. Type A are characterised by no skin involvement (the skin can be freely moved over the nodules). If the underlying tissues are not involved it is classified as nodular type A1 and if there is deep tissue involvement it is named A2 (Knottenbelt, 2005a). Nodular type B sarcoids are nodular sarcoids with visible and palpable skin involvement. When they have defined margins they are classified as B1 and if they have ill-defined margins they are named B2. Fibroblastic tumours have a 'fleshy' appearance (see Figure 1.2, Image 5). They are an aggressive form of sarcoids and have a tendency to ulcerate and become infected. Fibroblastic pedunculated sarcoids have a clear margin and a 'neck'

Figure 1.2. Examples of clinical types of equine sarcoids. 1. Occult. 2. Verrucose. 3. Nodular type A. 4. Nodular type B. 5. Fibroblastic pedunculated. 6. Mixed (Knottenbelt, 2005b)

that separates the lesion from healthy skin. Depending on deep tissue involvement, fibroblastic pedunculated sarcoids are classified as type 1a with no deep tissue involvement, and type 1b with deep tissue involvement. Fibroblastic sessile sarcoids on the other hand have ill-define margins, a broader base and there is always involvement of underlying tissue. It can become difficult to macroscopically differentiate any of the fibroblastic sarcoids from fibrosarcoma and squamous cell carcinoma in which case biopsy will be recommended. Other skin diseases that fibroblastic sarcoids may be mistaken with are exuberant

granulation tissue ('proud flesh'), a common complication in horses in wound healing. Mixed sarcoids are very variable in appearance and they can combine two or more of the forms described above. Malevolent equine sarcoids sometimes called malignant, are rarely described in the literature. The tumour infiltrates the lymphatic system and the progression is fast. It affects extensive areas and it may comprise different clinical types of ES.

1.1.3.2. Histopathology

Although histological findings are inconsistent, the microscopic appearance that defines equine sarcoids is characterized with acanthosis (epidermal hyperplasia), hyperkeratosis, *rete peg* formation (formations of superficial epidermal layers that run perpendicular to the skin surface, a normal finding in certain mucosae but it is abnormal in skin), picket fence fibroblast formation (fibroblasts sit perpendicular to the basal membrane) and proliferation of fibroblast-like cells (fusiform or spindle-shape fibroblast that form whorls) (see Figure 1.3) (Martens *et al.*, 2000). The epidermal surface may be normal, thinner or have been ulcerated (Marti *et al.*, 1993; Martens *et al.*, 2000).

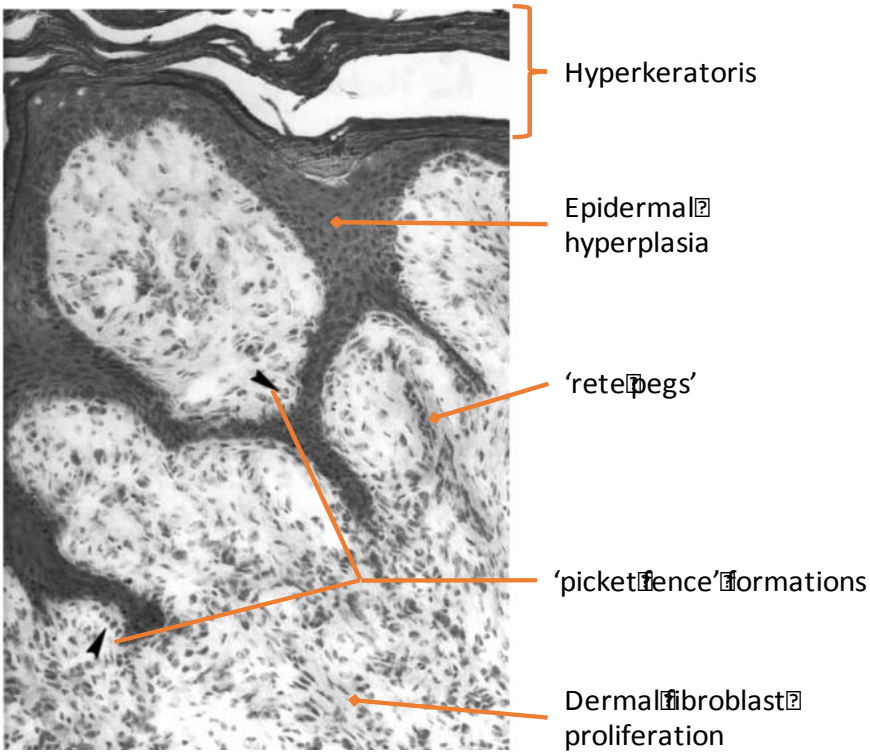


Figure 1.3. 160x microscopic amplification of a haematoxylin-eosin histopathology cut from an equine sarcoid, where all the most common features can be identified (Martens *et al.*, 2000)

1.1.3.3. Diagnostic test

Historically, clinical appearance and histological findings of sarcoids have been the tools to diagnose ES. For further details on histology the reader is advised to refer to section 1.1.3.2.

However, histology examination is not exempt of risk. Firstly, collecting biopsy specimens can be counterproductive for the patient, as the procedure can aggravate and worsen the progression of the tumour (Roberts, 1970; Martens *et al.*, 2001c). Secondly, the microscopic appearance of the lesions may be commonly confused with fibroma or fibrosarcoma and therefore the examination of the samples requires an experienced pathologist (Weiss, 1974; Taylor & Haldorson, 2013). Moreover, there is great variability in the histopathological signs and to date there are no pathognomonic features in equine sarcoids (Knottenbelt & Kelly, 2000). In a study based on 50 samples from five clinical types, Martens *et al.* (2000) identified increased proliferation of fibroblastic-like cells in all the tumours, although the occurrence of the rest of the histopathological features were highly variable. Differential diagnosis to consider when observing lesions compatible with sarcoids include excessive granulation tissue, granuloma, other papillomavirus lesions, fibroma/fibrosarcoma, cutaneous lymphoma, squamous cell carcinoma and, less commonly, habronema lesions, mast cell tumours, melanomas and staphylococcal folliculitis (Knottenbelt, 2005b).

Although the majority of sarcoid diagnoses are based upon clinical examination and histopathology, detection of BPV-1/-2 genomic DNA and viral expression in the masses can contribute to confirm the diagnosis of ES (Nasir & Reid, 1999; Carr *et al.*, 2001a; Nasir & Campo, 2008).

1.1.3.4. Treatment and prophylactic methods for equine sarcoids

ES are not metastatic but often show aggressive and infiltrative growth, often recur following ineffective treatment and animals with isolated masses frequently progress to develop multiple tumours (Goodrich *et al.*, 1998; Knottenbelt, 2005a). Spontaneous regression is only rarely observed (Brostrom, 1995; Knottenbelt & Kelly, 2000). Most lesions tend to develop to a more

aggressive type with time, especially when exposed to trauma (Hamann & Grabner, 2005).

No therapy exists that is universally effective for all cases of equine sarcoids. Treatment of sarcoids consists of removing the tumours through surgical excision (or with the use of a laser, liquid nitrogen or burning agents) followed or combined with cryotherapy, chemotherapy in the form of cysplatin beads or injections, intralesional fluoracil injection (5-FU), 'Liverpool cream' (marketed as AW3/4 ludes), immunotherapy (such as BCG or Imiquinol), local radiotherapy (iridium) and more recently, antiviral agents (Taylor & Haldorson, 2013). In a small percentage of cases, tumours spontaneously disappear and this may lead to the belief that the treatments were successful (Brostrom, 1995; Martens *et al.*, 2001b). It is not possible to compare post-treatment recurrence rate from the data available in the literature due to the bias in the population selected, not enough follow-up time to ensure no further development of the tumour, the differences in the therapeutic techniques, or in some studies the lack of confirmation of the diagnosis using histopathology. Recurrence rates are quoted from 5-64% (Tarwid, 1985; Knottenbelt *et al.*, 1995; Knottenbelt & Kelly, 2000; Martens *et al.*, 2001b). Vaccination has been attempted with variable success (Ashrafi *et al.*, 2008; Rothacker *et al.*, 2015), and alternative genetic therapy using knock-down of a BPV-1 oncoprotein E6 has shown promising results *in vitro* (Yuan *et al.*, 2011b). However to date, no preventative tools are available for equine clinicians and horse owners.

1.1.4. Equine sarcoid pathology

1.1.4.1. Aetiology

Jackson (1936) was the first to postulate that equine sarcoids may be caused by an infectious agent, based on clinical observations and pattern of spread of sarcoids in South Africa. A year after this report, the link between cancer and papillomavirus was established in rabbits (Rous & Beard, 1935; Jackson, 1936). A series of epidemiological and transmission studies followed, confirming autologous and homologous transmission of equine sarcoids giving strength to the theory that a transmissible agent was involved in the disease (Olson & Cook, 1951; Ragland *et al.*, 1966; Voss, 1969; Ragland *et al.*, 1970b). The transmission

studies used inoculation of extracts from bovine papillomas, extracts from naturally occurring sarcoids (NOS) and transplantation of excised tumours from sarcoid-bearing horses to healthy horses and donkeys. Unlike NOS, experimentally induced lesions were self-limiting, which indicates that additional factors are likely to contribute to the progression of the disease. Just a few years before the development of the Polymerase Chain Reaction (PCR), Lancaster *et al.* (1977) used DNA hybridisation to demonstrate the presence of bovine papillomavirus type 1 and type 2 genomes (BPV -1/-2) in NOS and artificially induced equine sarcoids. The advent of PCR led to several studies confirming the involvement of BPV-1/-2 in equine sarcoids and at present, there is strong evidence to suggest that Bovine papillomavirus types 1 and 2 (BPV-1/-2) are causally associated with in the disease (Angelos *et al.*, 1991; Chambers *et al.*, 2003a; Nasir & Campo, 2008; Nasir & Brandt, 2013). More recently the involvement of a third papillomavirus, BPV-13, has been documented in NOS (Lunardi *et al.*, 2013a). Whereas BPV-1 is most commonly found in ES in Europe and Australia, BPV-2 has been found to be more prevalent in ES in U.S.A. and Canada and BPV-13 in Brazil (Otten *et al.*, 1993; Bloch *et al.*, 1994; Carr *et al.*, 2001a; Bogaert *et al.*, 2007; Borzachiello *et al.*, 2008; Wobeser *et al.*, 2010; Lunardi *et al.*, 2013b).

There is an evident link between the presence of BPV-1/-2 and the development of ES but unfortunately many aspects of the disease remain poorly understood. *In vitro* experiments have proved that BPV-1 can transform equine primary fibroblasts to a phenotype seen in NOS (Yuan *et al.*, 2008a). BPV viral DNA is consistently present in the tumours (Amtmann *et al.*, 1980; Martens *et al.*, 2001c; Chambers *et al.*, 2003b; Bogaert *et al.*, 2007; Yuan *et al.*, 2007b; Borzachiello *et al.*, 2008) together with viral gene expression (Nasir & Reid, 1999; Carr *et al.*, 2001b; Chambers *et al.*, 2003b; Borzachiello *et al.*, 2008; Brandt *et al.*, 2011b; Wilson *et al.*, 2013). BPV-1/-2 DNA was undetected in other skin lesions including melanomas, papillomas and healthy skin (Carr *et al.*, 2001a; Bogaert *et al.*, 2007; Gaynor *et al.*, 2016) but present in equine inflammatory skin disease (Yuan *et al.*, 2007c; Wobeser *et al.*, 2012). Nevertheless, BPV-1/-2 DNA was detected in one case of squamous cell carcinoma (SCC) (Kainzbauer *et al.*, 2012) and hoof canker (Brandt *et al.*, 2011a). Only one report has proven the existence of BPV-1/-2 DNA in skin swabs

of a small group of horses free of sarcoids but in close contact with sarcoid-bearing animals (Bogaert *et al.*, 2005). However, long-term epidemiological data from the latter study were never reported. Presence of BPV-1 DNA in visibly healthy skin has also been reported in the margins of ES in sarcoid-bearing horses, up to 16mm from the margins of the tumours (Martens *et al.*, 2001a).

In sarcoids, viral DNA is detected largely in fibroblasts and is present at variable viral loads – reported from 2 copies per cell to several hundreds (Amtmann *et al.*, 1980; Yuan *et al.*, 2007b; Haralambus *et al.*, 2010; Brandt *et al.*, 2011b). One study identified viral genomes/per unit mass of tumour as a marker for severity of ES (severity was measured as rapid growth rate, presence of multiple tumours and high recurrence when treated) but not as a marker for clinical types (Bogaert *et al.*, 2007). When compared to naturally occurring tumours in cattle due to BPV-2, the number of BPV-1/-2 genomes per cell is substantially higher in equine sarcoids (Yuan *et al.*, 2007c).

Bovine papillomaviruses host range, classification and viral components

Bovine papillomaviruses (BPVs) belong to a heterogeneous family of non-enveloped circular double stranded DNA viruses, named *Papillomaviridae* (Bernard, 2006). In general papillomaviruses (PVs) are strictly species-specific (Smith *et al.*, 1984) and they affect only amniotes as they have been identified in mammals, birds and reptiles but not amphibians (Rector & Van Ranst, 2013).

The natural host of BPVs is cattle, in which the virus can cause benign tumours (papillomas and fibropapillomas) that more often than not regress, although occasionally they lead to malignant transformation affecting the gastrointestinal tract, udder, penis and bladder (Goodrich *et al.*, 1998; Chambers *et al.*, 2003b; de Villiers *et al.*, 2004; Silvestre *et al.*, 2009; Pangty *et al.*, 2010; Maiolino *et al.*, 2013; Roperto *et al.*, 2013). Three types of BPVs (BPV-1, -2 and -13) have been linked to the development of equine sarcoids therefore equine sarcoids are considered the first case of cross-species infection by PVs (Otten *et al.*, 1993; Nasir & Reid, 1999; Campo, 2003; Wobeser *et al.*, 2010; Lunardi *et al.*, 2013b).

Table 1.1. Summary of the 15 BPV types and main differences between the genotypes grouped in four genera

Genera	**DeltaPVs**	**EpsilonPVs**	**XiPVs**	**DyoxyPVs**
BPV types	BPV-1 BPV-2 BPV-13 BPV-14	BPV-5 BPV-8	BPV-3 BPV-4 BPV-6 BPV-9 BPV-10 BPV-11 BPV-12 BPV-15	BPV-7
Target cells	Fibroblasts	Superficial layers of the epithelium as well as fibroblasts	Superficial layers of the epithelium	Superficial layers of the epithelium
Type of lesion	Fibropapillomas	Epithelial lesions and fibropapillomas	Epithelial lesions	Epithelial lesions
Genome length	≈ 8 Kbp	≈ 8 Kbp	≈ 7.3 Kbp	≈ 7.3 Kbp
Other features	12 E2* protein binding sites in LCR**	10 E2* protein binding sites in LCR**	4 E2* protein binding sites in LCR** and no E6***	4 E2* protein binding sites in LCR**

***E2**= one of the early proteins
****LCR**=long control region of the BPVs genome
*****E6**= one of the early proteins with oncogenic function

All papillomaviruses are classified by nucleotide sequence of the major capsid protein L1 open reading frame (de Villiers *et al.*, 2004). To date over 316 distinct types of PVs have been described and they are grouped into 38 genera (Van Doorslaer *et al.*, 2013). The 15 types of BPVs characterised up to the present time (last update from 23/05/2016), are classified within four of these genera: Delta-, Epsilon-, Xi- and Dyoxipapillomaviruses. Details of the BPV types and main differences between genera are displayed in Table 1.1 (Bernard, 2006; Bernard *et al.*, 2010; Van Doorslaer, 2013).

The genome of BPV is approximately 8 kilobasepairs (Kbp) and it is enclosed in an icosahedral capsid of 55-60 nm in diameter (Campo, 2006). The genome is a double stranded circular DNA associated with histones and it is divided into 3 canonical regions: the early genes, the late genes and the long control region (LCR). A schematic representation of the layout of the BPV-1 genome is displayed in Figure 1.4.

The early genes encode non-structural regulatory proteins, which are essential for the life cycle, as well as oncoproteins (E5, E6 and E7). The early proteins participate in the modulation of the immune response and host cell growth, viral replication and regulation of transcription. Some of the early proteins (oncoproteins) are responsible for the oncogenic roles of BPVs (Bergman *et al.*, 1988; Chambers *et al.*, 2003b; Campo, 2006; Munday, 2014).

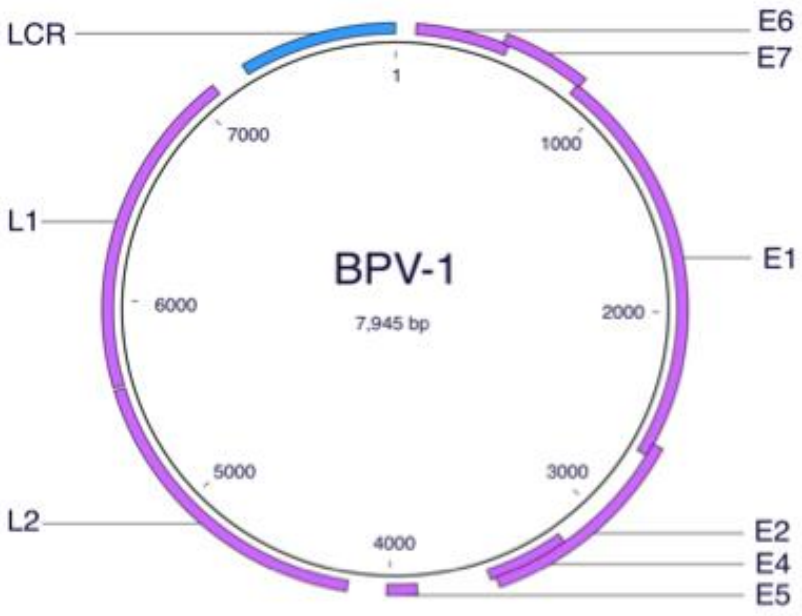


Figure 1.4. Diagrammatic genomic organisation of BPV-1. The circular genome of BPV-1 is 8Kbp long and it encodes six early proteins (E1, E2, E4, E5, E6 and E7) and two late proteins (L1 and L2), together with a long control region (LCR)

The late genes encode structural proteins for the capsid (L1 and L2) and the LCR is a non-coding region that contains the sequences of transcription promoters and enhancers, the origin of DNA replication and binding sites for cellular transcription factors (Spalholz *et al.*, 1987; Chambers *et al.*, 2003b; Nasir *et al.*, 2007).

BPVs life cycle

The viral cycle of BPVs, as for other PVs, is intimately linked to the differentiation of the epithelium and the virus uses the cell machinery for its

viral replication. A schematic representation of the known stages of the BPV-1 life cycle is displayed in Figure 1.5.

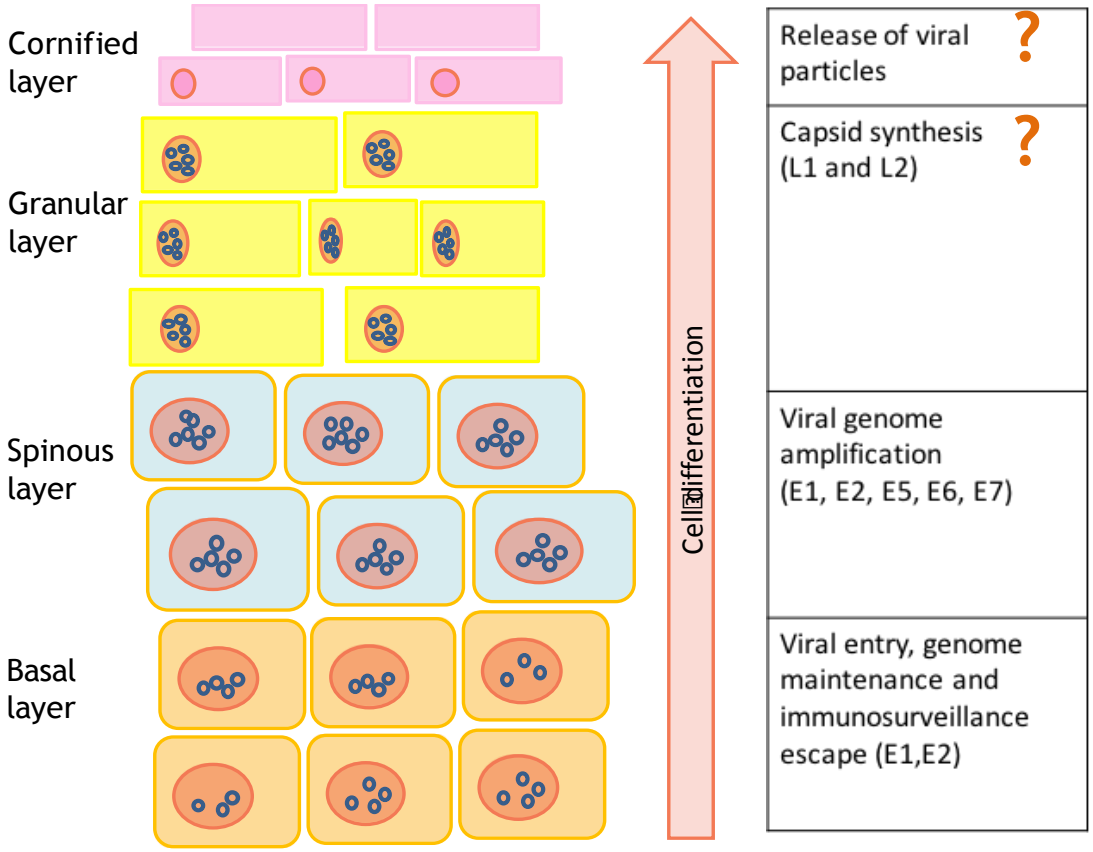


Figure 1.5. Schematic representation of the known stages of BPV-1 life cycle in ES. Stages in the upper layers such as capsid synthesis and viral particle formation have not been demonstrated in ES yet although expression of the late proteins and virus-like particles have been recently reported (Brandt *et al.*, 2008b)

It is hypothesised that in equine sarcoids, BPV-1 and -2 access basal layers of the stratified epithelium (basal keratinocyte) by microtrauma on skin/mucosa. Interestingly there is recent evidence of the DeltaPV BPV -2 infecting and replicating in transitional epithelium, placenta and lymphocytes in cows suffering from BPV-2 induced bladder cancer, and BPV-1 and -2 have been detected in white blood cells of horses with and without BPV-1/-2 induced skin lesions, suggesting that in Delta BPVs, there may be the possibility of vertical transmission as well as entry through the epithelium (Brandt *et al.*, 2008a; Hartl *et al.*, 2011; Roperto *et al.*, 2012; Silva *et al.*, 2013b). It is thought that BPVs, similar to HPVs, gain access to the intracellular space through L1 protein interactions with extracellular sulphated sugars (Joyce *et al.*, 1999). In other PVs, it is suggested that the virus is then enclosed in an endosome by interaction

with alpha-6 beta-4 integrin receptor (McMillan *et al.*, 1999) and the capsid protein L2 initiates the liberation of the genome from the endosome and entrance into the nuclear space (Kamper *et al.*, 2006).

BPVs form paracrystaline arrays in the nucleus of infected cells, and the genome is maintained at a low level (10-200 copies) as an extrachromosomal element (episomally) using the host machinery, as PVs do not encode polymerases (Park *et al.*, 1994). E1 and E2 proteins are essential during this phase. Basal cell replication maintains the infection but the virus needs to ascend to superficial layers to complete the life cycle. As infected cells from basal layers enter cellular division, viral DNA is evenly distributed between the two descendent cells and linked to the host chromosomes by the protein PV E2 (Baxter *et al.*, 2005); whilst one cell migrates to more suprabasal layers (spinous and granular) to initiate differentiation, the other cell containing the viral genome remains in the basal layer providing a reservoir of the viral genome. Infected basal and suprabasal cells express viral early proteins in the cytoplasm. Viral genome replication occurs in the suprabasal layers, and these cells in the absence of the virus are non-proliferative. Predominantly, the expression of E6 and E7 in HPVs and E5 in delta BPVs dysregulate the cell cycle, forcing differentiating cells into S phase, and allowing viral genome amplification in cells that normally would have exited the cell cycle (von Knebel Doeberitz, 2002; Venuti *et al.*, 2011; Doorbar *et al.*, 2012). PV oncoproteins can also activate a series of pathways that lead to uncontrolled proliferation, loss of differentiation of infected cells and loss of cell growth inhibition, and this is linked to accumulation of cellular mutations (persistency of S phase during mitosis) (Moody & Laimins, 2010).

In the PV life cycle in the natural host, expression of the late proteins to form the capsid is confined to the most superficial layers of the epithelium. Virus particles are then released when cells die. In equine sarcoids there is no evidence of mature virus being released, as infectious virions have not been isolated yet within equine sarcoids (Amtmann *et al.*, 1980). However, the presence of virus-like-particles (BPV-1 genome associated with L1 capsid formations) has been reported in a small group of ES using immunocapture PCR, a very sensitive technique (Brandt *et al.*, 2008b). This is the first evidence for the presence of precursor viral particles in equine sarcoids. There may be two reasons for the lack of infectious virions in equine sarcoids. Firstly, the

complexes reported by Brandt *et al.* (2008b) may be produced at a low level. Secondly, as horses are not the natural host of BPV-1/-2/-13 but non-permissive hosts, equine sarcoids may represent an abortive form of BPV infection in horses (Campo, 2003; Chambers *et al.*, 2003b).

In equine sarcoids, BPV-1 was initially thought to be restricted to equine fibroblasts of the dermal-epidermal junction (Teifke *et al.*, 1994). BPV-1 genomes have been recently demonstrated in both dermis and epidermis – including keratinocytes - of naturally occurring equine sarcoids, suggesting a more complex pathogenesis (Bogaert *et al.*, 2010; Brandt *et al.*, 2011b; Wobeser *et al.*, 2012).

BPV-1/-2 transmission in equine sarcoids

The transmission route of equine sarcoids is not yet clear. Direct transmission has been suggested between cattle or affected equids to other equids free of sarcoids (Jackson, 1936). This has been demonstrated in donkeys where close contact can result in the development of ES (Reid *et al.*, 1994). The presence of viral DNA in the epidermis of horses with inflammatory conditions suggests the possibility of transmission between individuals via direct contact (Wobeser *et al.*, 2012). Indirect transmission through a vector or fomites seems plausible. BPV-1/-2 may spread in insect vectors where it has been found (Kemp-Symonds, 2000; Finlay *et al.*, 2009). This hypothesis is also supported by clinical and epidemiological data. Some horses develop multiple tumours, ES can occur in intact skin and horses that have not been in contact with cattle can develop ES (Voss, 1969; Reid *et al.*, 1994; Torrontegui & Reid, 1994; Knottenbelt *et al.*, 1995). These facts suggest that vectors may play an important role in the disease. Indirect transmission through fomites such as tack equipment or brushes is another possibility as the papillomaviruses resist desiccation for days maintaining infectivity (Roden *et al.*, 1997). Vertical transmission has not been demonstrated in ES although BPV-1 DNA and viral expression has been reported in white blood cells of horses with experimentally induced ES (Brandt *et al.*, 2008a; Hartl *et al.*, 2011). In cattle affected by BPV-2 bladder cancer, presence of the virus in placenta suggests the possibility of vertical transmission (Roperto *et al.*, 2012).

BPV-1/-2 oncoproteins and their role in tumoural development

As in high risk-HPV (HR-HPV) induced cancers, BPV -1/-2 presence is essential for the development of ES but infection alone is not sufficient for tumour production (Bogaert *et al.*, 2005; Yuan *et al.*, 2007c; Bogaert *et al.*, 2008). Bradford Hill criteria for viruses as causality of cancer is a widely used guide that utilises certain clinical and epidemiological data to establish the link between viruses and human cancers (Hill, 1965); unfortunately such data is not available for equine sarcoids.

In HR-HPV induced cancers the virus is thought to promote the oncogenic process through several mechanisms such as allowing accumulation of genetic mutations and expressing proteins that alter the cell cycle and promote cell proliferation (Moody & Laimins, 2010). In 50% of HR-HPV induced cancers, tumoural transformation is linked to integration of the HPV genome into the host genome at points of genetic instability and it is linked to overexpression of oncoproteins E6 and E7 (Moody & Laimins, 2010). This is not the case for BPVs, which maintain their genomes episomally and in which the oncogenic effect is attributed to three early oncoproteins (E5, E6 and E7) in both ruminant cancers and equine sarcoids (Yuan *et al.*, 2011b; Corteggio *et al.*, 2013). In equine sarcoids, E5 and E6 have been proven to be the major transforming oncoproteins and E7 has a contributing oncogenic effect (Yuan *et al.*, 2011b).

BPV-1/-2 E5 oncoprotein

Compared to high risk HPV-16 E5, BPV-1 E5 is small (83 *vs.* 44 amino acids long respectively), and by itself BPV-1 E5 has a higher transforming *in vitro* activity, as it is sufficient to fully transform mouse, primary human fibroblasts and equine fibroblasts (Marchetti *et al.*, 2002; Yuan *et al.*, 2011b). BPV-1 E5 has two main parts: an amino-terminal domain, with a strongly hydrophobic leucine-rich membrane-spanning region and a small hydrophilic single amino acid (glutamine) residue in position 17; and a hydrophilic carboxyl-terminal domain (Venuti *et al.*, 2011; Corteggio *et al.*, 2013). The hydrophilic residues at position 17 of the amino-terminal domain and the carboxy-terminal tail are thought to be essential for the transforming effect of the BPV-1 E5 protein as they facilitate dimerization of the E5 protein and facilitate interaction with ATPase 16 K

subunit C and beta-type platelet-derived growth factor receptor (PDGFß-R) (Horwitz *et al.*, 1988; Goldstein *et al.*, 1992; Klein *et al.*, 1999). Seven variants in the amino-acid sequence of E5 in equine sarcoids have been identified when compared to BPV-1 E5 from bovine fibropapillomas. However, the glutamine in position 17 was found unchanged across five equine sarcoid BPV-1 E5 variants (Chambers *et al.*, 2003a).

BPV-1 E5 is found in the Golgi apparatus, endoplasmic reticulum and inner part of the cytoplasmic membrane of basal cells and superficial granular cells in bovine fibropapillomas and BPV-1 transformed bovine primary fibroblasts (Ashrafi *et al.*, 2002; Venuti *et al.*, 2011; Corteggio *et al.*, 2013; Bocaneti *et al.*, 2016). The expression of the E5 protein in both basal cells and most superficial granular layers, including keratinocytes in bovine fibropapillomas suggests a role in both early stages and late stages of the viral infection (Corteggio *et al.*, 2013). In the natural host, BPV-1 E5 activates PDGFß-R and subsequently upregulates the phosphatidylinositol – 3 - (PI3K)–protein kinase B (known as AKT)-cyclin D pathway, leading to cell cycle deregulation (Araibi *et al.*, 2004; Borzachiello *et al.*, 2008; Corteggio *et al.*, 2010). PDGFß-R recruits growth-factor-receptor-bound protein 2 (GRB2) and 'son of sevenless' gene protein (SOS) leading to Ras activation but not to phosphorylation of the two mitogen-activated-protein kinases extracellular-signal - regulated (Erk) and microtubule associated protein (Mek or MAPK2) kinases (Corteggio *et al.*, 2012). Based on this, the PI3K-AKT-cyclin D rather than Ras-Erk pathway seems to be the important pathway in BPV urothelial carcinogenesis. BPV-1 E5 activates the proto-oncogene tyrosine-protein kinase (c-src) in bovine bladder carcinoma and it is hypothesised to be involved in the loss contact inhibition due to downregulation of gap junction communication through two mechanisms, deregulation of conexin26 and binding to the ductin section of the 16KDa cellular protein (Borzacchiello & Roperto, 2008; Silva *et al.*, 2013a; Gil da Costa & Medeiros, 2014).

Expression of BPV-1 E5 in equine sarcoids has been demonstrated in both dermis and epidermis (Brandt *et al.*, 2011b). At a cellular level the oncoprotein of BPV-1/-2 was found in the cytoplasm adjacent to the nucleus (Borzachiello *et al.*, 2008; Altamura *et al.*, 2013). The best studied tumorigenic effects of BPV-1 E5 in equine sarcoids are related to the PDGFß-R and the equine major

histocompatibility complex (MHC) class I. BPV-1 and -2 E5 induce tumorigenic transformation to cell cultures by strongly and specifically binding to its cellular target, PDGFß-R in a ligand-independent manner (Borzachiello *et al.*, 2008). Increased amount of activated PDGFß-R has been found in equine sarcoids and human fibroblasts expressing BPV-1 E5 oncoprotein, to bind a regulatory subunit of PI3-K together with increased expression of AKT and cyclin D3 which results in mitogenic signalling activation (Borzachiello *et al.*, 2008). Activation of PDGF-ß receptors of the dermal fibroblasts results in mitogenesis in nearby epithelial cells (Carr *et al.*, 2001b), contributing to the pseudoepitheliomatous hyperplasia that is typically observed in BPV infected lesions. MHC class I is an important part of the immunosurveillance and it is responsible for antigen presentation to T-cells (Garcia-Lora *et al.*, 2003). BPV-1/-2 E5 downregulates MHC class I by reducing the transcription and promoting its degradation (Ashrafi *et al.*, 2002; Yuan *et al.*, 2008b). Downregulation of MHC class I through intracellular retention leads to immune evasion as the molecules are not presented on the cell surface (Chambers *et al.*, 2003b). Intracellular retention of MHC class I is thought to occur as a consequence of BPV-1 E5 binding to subunit C of the 16KDa cellular protein and affecting intracellular pH leading to acidification (Borzacchiello & Roperto, 2008). Decrease in MHC class I has been reported in BPV bovine cancer and ES (Ashrafi *et al.*, 2002; Marchetti *et al.*, 2002; Yuan *et al.*, 2008b). Other possible effects demonstrated of BPV-1 E5 the activation of p38 and MAPK that leads to increased cell proliferation, and increase of MMP-1 (Yuan *et al.*, 2010b; Yuan *et al.*, 2011c). A summary of the known oncogenic mechanisms of BPV-1 E5 in ES is displayed in Figure 1.6.

BPV-1/-2 oncoproteins E6 and E7

BPV-1 E6 and E7 are also small proteins (18 and 13 KDa respectively) located in the nucleus and in the cytoplasmic compartment and their oncomechanisms in bovine and equine tumorigenesis are not fully understood (Teifke *et al.*, 1994; Borzachiello *et al.*, 2008). In high risk human papillomavirus (HR-HPV), presence of E6 and E7 is essential for tumoural development but not sufficient, as there are patients expressing these viral proteins that do not develop cancer (Moody & Laimins, 2010). HR-HPV E6 and E7 affect a series of cellular pathways that lead to DNA damage and mutation resulting in a cancerous phenotype. HR-HPV E7 protein binds and downregulates retinoblastoma protein (pRB). pRB is a tumour

suppressor and it regulates E2F promoters, a series of genes that code for transcription factors that intervene in cell cycle regulation, cell differentiation, mitosis and apoptosis. HR-HPV E7 also frees up E2F promoters by disrupting pRB-E2F complexes. The consequence of alteration of these pathways is the disruption of the cell cycle. A second pathway proven to be associated with tumoural transformation in HR-HPV is the tumour protein p53 pathway. HPV E6 and E7 affect p53 in opposite ways. p53 repairs DNA, is a tumour suppressor and regulates apoptosis amongst other functions. E6 and E7 activity on the p53 pathway makes the cells virus permissive, allowing viral genome replication and, as a consequence of alteration of the cell cycle, the host DNA accumulates mutations (Moody & Laimins, 2010). HR-HPV E6 also activates transcription of telomerase reverse transcriptase (TERT), an essential step for cell immortalisation.

BPV-1 E6 interaction with Crebb binding protein (CBP) and protein p300 leads to impairment of the function of tumour suppressor p53 (Zimmermann *et al.*, 2000). Binding of BPV-1 E6 with paxillin leads to alterations in the actin cytoskeleton that are intrinsically related to tumorigenesis and BPV-1 E6 has recently been reported to bind mastermind-like protein1 (MAML1), delaying keratinocyte differentiation by impairment of the Notch signalling pathway and subsequently gene transcription (Tong & Howley, 1997; Brimer *et al.*, 2012; Brimer *et al.*, 2014). Little is known about the mechanisms of BPV-1 E6 in ES. Expression of p53 was found consistent in 4 different clinical types of ES and mutation of p53 seems to be an uncommon event (Nasir *et al.*, 1999; Bogaert *et al.*, 2007). At a cellular level, presence of the virus triggered p53 mislocation by an unknown mechanism (Martens *et al.*, 2000; Yuan *et al.*, 2008b). Even though distribution of p53 is abnormal in sarcoid fibroblasts, BPV-1 E6 in ES is thought to promote tumorigenesis by p53 independent mechanisms.

BPV-1 E7 in bovine fibropapillomas is expressed in spinous layers and in basal layers (Bohl *et al.*, 2001). Co-localization of E7 with BPV-1 E5 in basal layers suggests that E7 may have a role in cell proliferation control (Bohl *et al.*, 2001). BPV-1 E7 expression in suprabasal layers is linked in bovine fibropapillomas to the over-expression of beta-1 integrin and proliferating cell nuclear antigen (PCNA) and it is thought to relate to cell proliferation and cell-cycle dysregulation (Cooper *et al.*, 2006). BPV-1 E7 sensitised cells to TNF induced

apoptosis (Liu *et al.*, 2000). Unlike high risk HPV E7, BPV-1/-2 E7 protein does not have a pRB binding domain but contributes to tumoural transformation in bovines by blocking pRB function by binding to a 600-kDa retinoblastoma protein associated factor (p600) leading to inhibition of anoikis (cancer cell apoptosis mechanism) (DeMasi *et al.*, 2005; DeMasi *et al.*, 2007).

BPV-1 E7 is found in the cytoplasm and nucleus in equine sarcoids (Borzachiello *et al.*, 2008). This protein is co-expressed with the aforementioned BPV-1 E5 suggesting a co-operative role of the E7 protein in tumoural transformation. In equine sarcoid tumours, BPV-1 E7 has been demonstrated to bind protein p600 therefore contributing to cell transformation by inhibiting anoikis (Corteggio *et al.*, 2011). It is through this mechanism that BPV-1 is thought to promote overexpression of matrix metalloproteinases (MMPs) contributing to invasiveness (Corteggio *et al.*, 2011; Yuan *et al.*, 2011a). Additionally, BPV-E6 is thought to contribute to MMPs upregulation in ES as both activate activator protein -1 (AP-1). Overexpression of MMPs is linked to alteration of mechanisms of immunity and inflammation, cell adhesion, motility, and integrity as well as cell cycle/cell proliferation and apoptosis (Yuan *et al.*, 2008b; Yuan *et al.*, 2011b). Another effect attributed to BPV-1 E7 (and shared with BPV-1 E2) is the decrease in expression of toll-like receptor type 4 (TLR4) in ES (Yuan *et al.*, 2010a). Oncogenic mechanisms of BPV-1 E6 and E7 in ES have been summarised in Figure 1.7.

Since many aspects of the tumoural transformation in equine sarcoids remain unknown, and the diagnostic, therapeutic and prophylactic tools currently available do not provide sufficient help to control or prevent the disease, there is a particular need for further understanding of the pathogenesis of equine sarcoids. To aid in the discovery of new diagnostic and therapeutic tools, it is essential to have knowledge of the molecular mechanisms by which the virus leads to the tumoural transformation. Recently, miRNAs, small RNA molecules that are known to control essential pathways within the cell, have been extensively studied in disease mechanisms including cancers and HPV induced disease. Their biogenesis, function and known role in cancer as well as current knowledge of the miRNAs in the horse will be fully described in the following section.

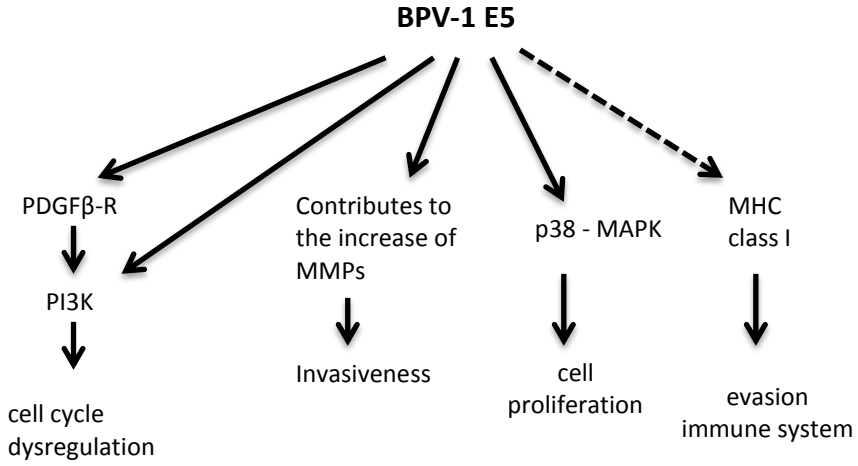


Figure 1.6. Summary of the oncogenic mechanisms of BPV-1 E5 demonstrated in ES. Solid arrows indicate upregulation or activation and dashed arrows indicate downregulation

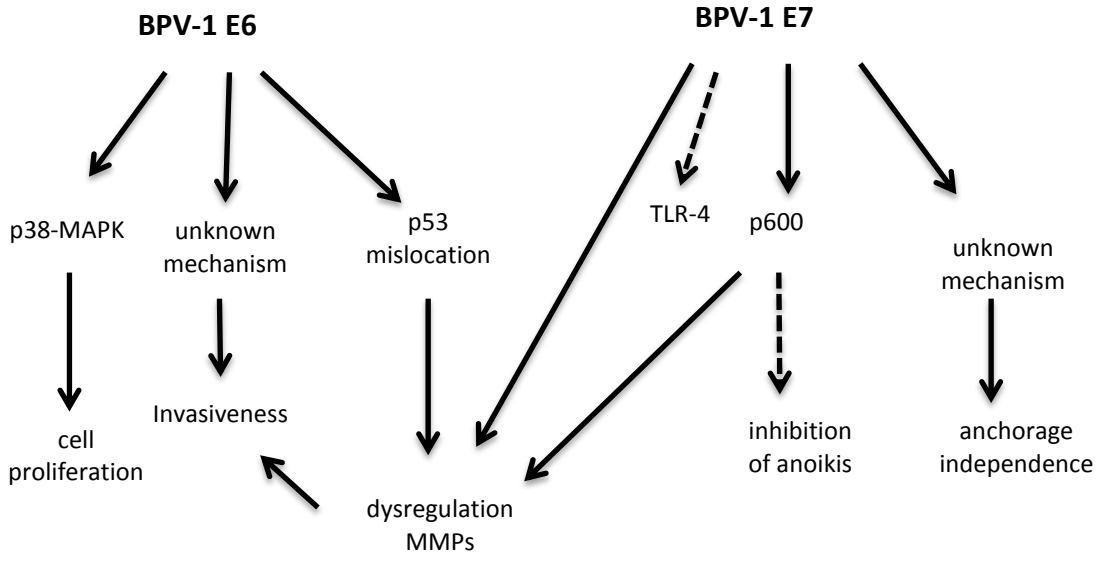


Figure 1.7. Summary of the oncogenic mechanism of BPV-1 E6 and E7 demonstrated in ES. Solid arrows indicate upregulation or activation and dashed arrows indicate downregulation

1.2. microRNAs (miRNAs)

1.2.1. Definition

microRNAs (miRNAs) are 18-23 nucleotide, single stranded RNA molecules that regulate gene expression at the post-transcriptional level (Ambros, 2004). miRNAs were first discovered in the free-living model nematode *Caenorhabditis elegans,* with the sequencing of miRNA lin-4 (Lee *et al.*, 1993). These molecules are now known to be responsible for regulation of the cell cycle, cell proliferation and differentiation, apoptosis and senescence in a wide range of organisms (Kanellopoulou *et al.*, 2005; Hermeking, 2007; Melton & Blelloch, 2010). Importantly, dysfunctional expression of miRNAs has been linked to malignancies (Calin *et al.*, 2002; Bhardwaj *et al.*, 2010; Zheng & Wang, 2011). The function of miRNAs is evolutionarily conserved and these molecules are receiving increasing attention as potential biomarkers for diagnosis, prognosis, molecular classification, monitoring progression and treatment for many diseases (Srivastava *et al.*, 2013). miRNAs can be expressed in a tissue- and species-specific manner and they are stable in body fluids, tissues and paraffin embedded tissue samples (Srivastava *et al.*, 2013).

1.2.2. Biogenesis and mechanism of action of miRNAs in eukaryotic cells

A summary of the biogenesis and mechanism of action of miRNAs is presented in Figure 1.8. miRNAs are expressed in a primary precursor form, referred to as a pri-miRNA, which is the transcription product of RNA polymerase II/III. These molecules, of hundreds to thousands of nucleotides in length, have a characteristic low energy fold-back hairpin secondary structure that lacks large internal loops of bulges (Cai *et al.*, 2004; Lee *et al.*, 2004). Primary miRNAs are processed into secondary precursors or pre-miRNAs of approximately 70 nucleotides in length by a class 2 ribonuclease III enzyme named Drosha, that is associated with the microprocessor complex subunit DiGeorge syndrome chromosomal (or critical) region 8 (DGCR8 protein). Pre-miRNAs are then exported out of the cell nucleus by Exportin-5 and subsequently processed by another RNAse III Dicer to form the characteristic miRNA hairpin duplex

structure. miRNA hairpins comprise two complementary miRNA mature sequences: the active miRNA (guide strand) and the complementary passenger strand historically referred to as the star sequence (miRNA*). The current annotation refers to these as the miR-5p and miR-3p referring to their location at the 5' or 3' end of the RNA transcript. The guide mature form is considered to be the dominant, active molecule; the star strand was thought to undergo degradation. Nevertheless, recent studies have demonstrated that actually the passenger strand (miRNA*) can have a function (Jagadeeswaran *et al.*, 2010). The ratio of mature to miRNA* abundance can vary between tissues and stages, a process referred as arm switching, which is thought to allow functional evolution of miRNAs (Griffiths-Jones *et al.*, 2011). miRNAs inhibit mRNA translation causing deadenylation and/or degradation of their target mRNAs (Filipowicz *et al.*, 2008). The guide miRNA strand is incorporated into the RNA-induced silencing complex (RISC) and directs the complex to mRNA target sequences, leading to translational repression. Specificity is achieved by imperfect base-pairing of the miRNA, most commonly to the 3'-untranslated region (UTR) of target mRNA sequences. There is evidence to suggest that 5'-UTR base pairing can also occur (Grey et al., 2010). The seed sequence of a miRNA, nucleotides from position 2-7/8 from the 5' end (Lewis *et al.*, 2005; Filipowicz *et al.*, 2008; Friedman *et al.*, 2009; Kim *et al.*, 2009) is the region thought to be crucial for complementary base-pairing with target mRNAs. Closely related microRNAs are grouped into families based on homology of the seed sequence, and miRNAs in the same family are believed to be functionally equivalent in certain situations. One mature miRNA can be encoded by multiple regions of the genome and hence derive from different precursors (Wang *et al.*, 2014a). One miRNA may target multiple mRNAs and conversely many mRNAs may be targeted by several miRNAs (Filipowicz *et al.*, 2008). These mechanisms enable microRNAs to 'switch on' and 'switch off' pathways involved in the cell cycle, proliferation, apoptosis, and senescence (Lu *et al.*, 2005; Rosenfeld *et al.*, 2008; Wang *et al.*, 2011; Wu *et al.*, 2014). These levels of complexity of miRNA-mRNA interactions make functional analysis of miRNAs challenging.

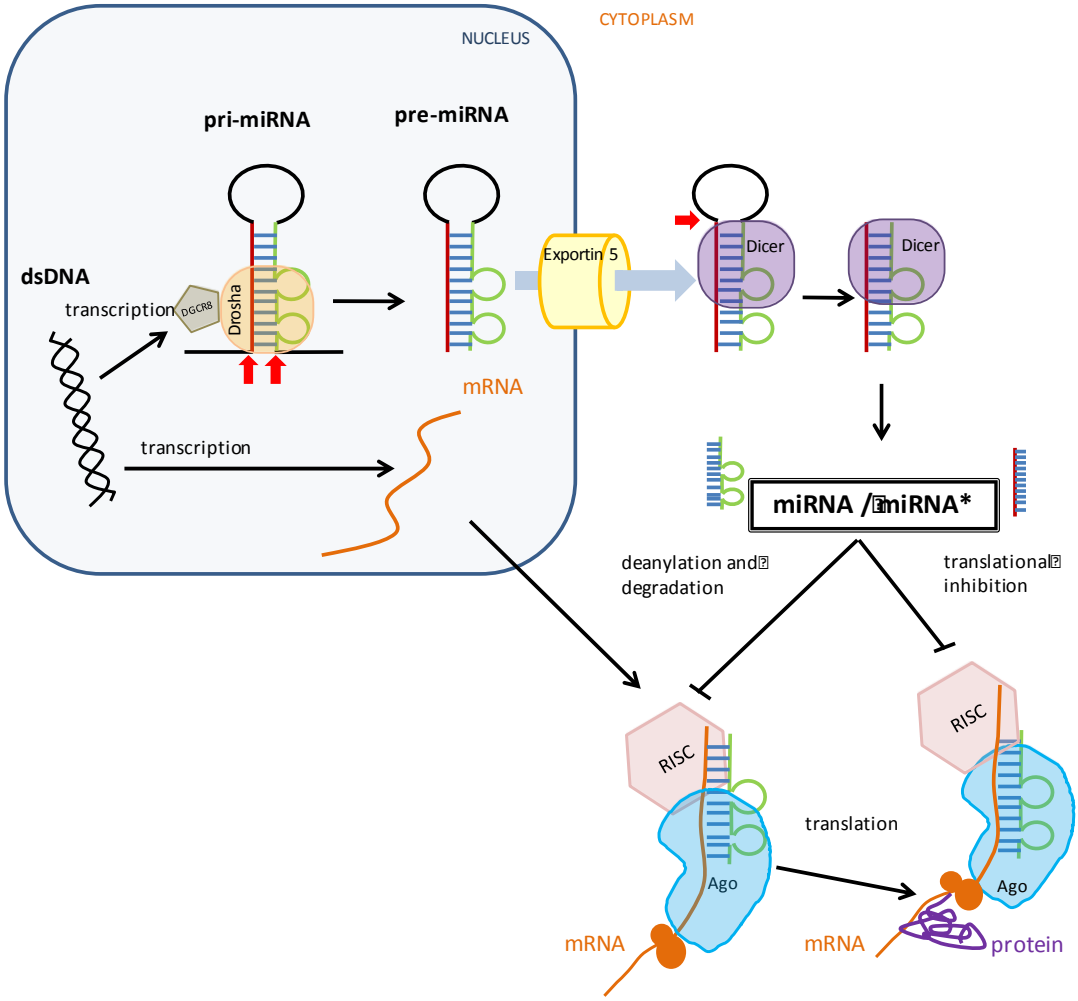


Figure 1.8. Schematic representation of biogenesis and mode of action of miRNAs. The first microRNA precursor (pri-miRNA) is transcribed in the nucleus and subsequently cut by a RNAse (Drosha) to form a 70-110 nucleotide long precursor (pre-miRNA). This precursor is exported to the cytoplasm and cut by another RNAse (Dicer) to produce the mature dominant microRNA and the miRNA*. By imperfect base pairing the active/dominant miRNA binds into the target mRNA as part of a protein complex named RISC (RNA-induced silencing complex), that contains proteins from the argonaute family (Ago)

1.2.3. miRNAs in cancer

Interest in miRNAs as important regulators of gene expression arose from their identification not only in nematodes but in the human genome (Giza *et al.*, 2014). Dysfunctional expression of microRNAs has been linked to multiple pathological mechanisms including cancer (Calin & Croce, 2006; Venugopal *et al.*, 2010; Kim *et al.*, 2011a; Baek *et al.*, 2015). Moreover, they are located in fragile regions of the genome and they have been shown to bind regions of the genome with chromosomal instability and/or amplification (Bhardwaj *et al.*,

2010; Liu, 2013). miRNAs involved in cancer are named oncomiRs (Lotterman *et al.*, 2008). By repressing certain cellular processes, some miRNAs can have an oncogenic effect or tumour suppressor effect (Huang *et al.*, 2015; Tutar *et al.*, 2015). miR-15 and miR-16 were the first two oncomiRs reported. They are encoded on human chromosome 13 and were found to be deleted in patients with B cell chronic lymphocytic leukaemia (Calin *et al.*, 2002). Since then, multiple studies have detected changes in miRNA expression linked to different types of cancer (Calin & Croce, 2006; Kim *et al.*, 2011a; Kim *et al.*, 2011b; Lajer *et al.*, 2012; Sharma *et al.*, 2014). miRNAs have been proposed as biomarkers for cancer, signatures for diagnosis and progression, and even as therapeutic tools (Heneghan *et al.*, 2010; Wittmann & Jäck, 2010; Srivastava *et al.*, 2013; Wang *et al.*, 2014b; Nagaraj *et al.*, 2015; Roth *et al.*, 2015).

One of the most commonly reported dysregulated group of mature miRNAs in cancer are members of the let-7 family, the sequence of which is perfectly conserved in a wide range of animal species, but not present in plants (Pasquinelli *et al.*, 2000; Hertel *et al.*, 2012). The let-7 family is involved in regulating cell differentiation and limiting cell proliferation and let-7 family miRNAs have been identified as important tumour suppressor genes (Pasquinelli *et al.*, 2000; Lee *et al.*, 2016). Members of the let-7 family are down-regulated in many tumours (Li *et al.*, 2011; Yu *et al.*, 2015; Hung *et al.*, 2016) and one of the main pathways disrupted in carcinogenesis, the Ras-mitogen activated protein kinase (Ras-MAPK) pathway, is directly regulated by the let-7 family (Johnson *et al.*, 2005; Masliah-Planchon *et al.*, 2015). Other examples of tumour suppressors are miR-15a/16, miR-26, miR-29/181, miR-39a, miR-122 miR-143/145, miR-150, miR-200b/c (Lotterman *et al.*, 2008; Liu, 2013; Soon & Kiaris, 2013).

Examples of miRNAs that promote cancer development are miR-21, cluster miR-17-92, miR-191, miR-155, miR-372 and miR-373 (Lotterman *et al.*, 2008; Heneghan *et al.*, 2010). It is important to acknowledge that due to the nature of the miRNA-target mRNA interactions, the same miRNA may function as a suppressor or as a tumour activator (Calin & Croce, 2006).

1.2.4. miRNAs in papillomavirus induced cancer

Scarce information is available on the changes in miRNA expression in papillomavirus cancer, except for HPV- induced cancer. The effects of miRNAs in high risk HPV cancer has been extensively studied and during the last couple of years comprehensive reviews have been published by different groups (Banno *et al.*, 2014; de Freitas *et al.*, 2014; Sharma *et al.*, 2014; Diaz-Gonzalez *et al.*, 2015; Wang *et al.*, 2016). The effect of HPV on miRNA dysregulation is discussed in more detail in Chapter 5. Although there is some variation in a small number of miRNAs, several studies have shown altered expression of miRNAs in HPV cervical cancer. To name a few, HPV-16 causes *in vivo* upregulation of miR-16/15b, miR-21 and downregulation of miR-34a, miR-143/145, miR-203 and miR-218 (Wang *et al.*, 2008; McBee *et al.*, 2011). These changes lead to an inability to stop cell proliferation and other mechanisms that trigger tumoural transformation (Zheng & Wang, 2011; Lajer *et al.*, 2012). Some of the miRNAs dysregulated in HPV induced cancer, such as miR-21, miR-143/145, miR-218 or miR-34a, have been suggested as biomarkers for the diagnosis of the disease and of disease progression (de Freitas *et al.*, 2014).

Viruses encode miRNAs and interestingly five potential miRNA candidates have been reported in HPV-16 located in the E5, E1 and L1 genes, and in the LCR region (Qian *et al.*, 2013). However, other groups have not reported any miRNAs encoded in HPV (Wang *et al.*, 2008).

1.2.5. miRNAs in the horse

The horse genome (*Equus caballus*) was first assembled and released seven years ago (EquCab2.0, GenBank assembly accession reference GCA_000002305.1) and since then a small number of studies have started to explore microRNA profiling in the horse. Since the first set of 354 mature microRNAs was predicted using an *in silico* method by Zhou *et al.* (2009), equine microRNA research has focused on disease-related experiments or breeding performance (Barrey *et al.*, 2010; da Silveira *et al.*, 2012; Buechli *et al.*, 2013; Das *et al.*, 2013; Donadeu & Schauer, 2013; da Silveira *et al.*, 2014; Desjardin *et al.*, 2014). Three experimental approaches have been used to date to identify potential equine microRNAs: *in*

silico prediction (Zhou *et al.*, 2009), microRNA quantitation using real-time quantitative reverse-transcription polymerase chain reaction (real-time qRT-PCR) either on a commercial platform or as individual tests (Barrey *et al.*, 2010; da Silveira *et al.*, 2012; Buechli *et al.*, 2013; Donadeu & Schauer, 2013; da Silveira *et al.*, 2014) and microRNA transcriptome analysis using high throughput sequencing (HTS also known as RNA-seq technology) (Desjardin *et al.*, 2014; Kim *et al.*, 2014; Platt *et al.*, 2014). The main findings of these studies are summarised in Table 1.2.

1.2.5.1. Predicted (in silico) microRNAs in the equine genome

Zhou *et al.* (2009) performed the first *in silico* integrated comparative analysis of equine miRNAs and predicted 354 mature miRNA sequences in the equine genome using bioinformatic tools. These miRNAs were mapped to 407 genes, located on all 32 chromosomes except ECA19 and ECA31. Out of the 354 mature microRNA sequences, 75 were grouped into 32 families. This was achieved by comparing the entire database of animal microRNAs (miRBase) to the most updated version of the genome of the horse (EquCab2.0) with bioinformatic tools. No experimental validation was performed in the study of Zhou et al. and the microRNA candidates were predicted using a series of assumptions:

- A total number of five nucleotide mismatches with microRNAs from other animal species was accepted.
- The search in Basic Local Alignment Search Tool (BLAST) was modified to analyse smaller fragments of RNA of seven nucleotides, rather than the default number (11 nucleotides length).
- Expectation value (e value) of ten was considered acceptable.
- Structural characteristics as well as minimal folding free energy were taken into account for the predictions (MFold and MiPred softwares).

Table 1.2. List of studies reporting equine miRNAs in the literature with a summary of the techniques and findings

Source	Species	Type of study	Number of horses / samples	Type of sample	Condition under study / method	Number of miRNAs of interest / discovered	Annotations of the miRNAs of interest
Zhou et al., 2009	Equine	in silico prediction	not applicable - in silico study	genomes from all animals and genome assembly of the horse (EquCab2)	Descriptive, comparative and predictive (integrated method)	354 mature miRNA predicted candidates - 407 genes	not applicable - in silico study
Barrey et al.,2010	Equine	real-time qRT-PCR (Taqman®) - human sequences	20 Cobs/French trotters including controls	muscle and blood	Recurrent exertional rhabdomyolysis (RER) and polysaccharide storage myopathy (PSSM) (muscular diseases)	10	miR-1, miR-23a, miR-30b, miR-133, miR-181, miR-188, miR-195, miR-206, miR-339, miR-375
da Silveira et al., 2012	Equine	real-time qRT-PCR (SYBR green) - human sequences (miRNome)	not specified	ovarian follicular fluid	Differences between old and young mares (breeding performance)	3	14-signature miRNA of 'young mares' and 5 of 'old mares' in exosomes. Differentially expressed: miR-181a, miR-375 and miR-513a-3p
Buechli et al., 2013	Equine	real-time qRT-PCR (Taqman®) - human sequences	not applicable - in vitro study	cells	Screening in chondrocytes (orthopaedic conditions)	1	miR-140
Das et al., 2013	Equine	RNA-seq, no validation reported	5 stallions	semen	Descriptive transcriptome semen in the horse (breeding performance)	82	miR-34B, miR-34C, miR-191, miR-223, miR-1248, and miR-1905C are the most expressed miRNAs
Donadeu and Schauer, 2013	Equine	real-time qRT-PCR (SYBR green) - human sequences	5 mares	follicular exosomes	Follicular growth (breeding performance)	9	miR-21, miR-23a, miR-145, miR-503, miR-224, miR-383, miR-378, miR-132 and miR-212

Table 1.2. (Continuation) List of studies reporting equine miRNAs in the literature with a summary of the techniques and findings

Source	Species	Type of study	Number of horses / samples	Type of sample	Condition under study / method	Number of miRNAs of interest / discovered	Annotations of the miRNAs of interest
Desjardin et al., 2014	Equine	RNA-seq and validation using real-time qRT-PCR (Taqman®)	3 10-month old foals with osteochondrosis and 3 10-moth old foals with no osteochondrosis	15 samples from bone and 15 samples from cartilage from each group	Osteochondrosis (orthopaedic diseases)	371 mature miRNA candidates* 299 suggested novel miRNAs**	DE of microRNAs in bone and OC was found in controls vs predisposed
Buza et al., 2014	Human, cow, chicken, pig, dog, equine	In silico prediction	not applicable - in silico study	not applicable	Comparative study	146 miRNAs present in the horse#	not specified
da Silveira et al., 2014	Equine	real-time qRT-PCR (SYBR green) - human sequences (miRNome)	12 mares	ovary	Different parts of the oestrus cycle (breeding performance	4	miR-27b, miR-372, miR-382 and miR-27b
Plat et al., 2014	Bat, dog, equine	RNA-seq and validation using real-time qRT-PCR (Taqman)	not specified	1 testicular tissue	Comparative study	not specified	not specified
Kim et al., 2014	Equine	RNA-seq, no validation reported	8 Thoroughbred horses	liver, colon and muscle	Descriptive	292 previously described *** 392 novel **	not applicable - transcriptome study

*some of the miRNA annotations are not in miRBase; **miRNA sequences not submitted to miRBase (02/2016); *** miRNAs have significant variation to the sequences in miRBase; #present in the horse and in humans are associated with disease in the database HDMACs (human disease microRNAs with animal counterpart)

1.2.5.2. Real-time quantitative reverse transcription PCR (qRT-PCR) quantification of miRNAs using human/mice sequences for the primers/probes.

Quantification of microRNAs using qRT-PCR has been achieved either with DNA binding dyes (SYBR green) or probe-based chemistries (Taqman®). SYBR green I dye binds to any double stranded DNA (dsDNA). TaqMan® probes contain a fluorescent reporter dye at the 5' end and a quencher dye at the 3' end. During the PCR reaction, the dye is separated from the quencher initiating fluorescence liberation. SYBR green quantification is less expensive, and less specific. The advantage of SYBR green is the melting curve of the PCR products that allows examination of any non-specific amplification or contamination. On the other hand, Taqman® technology offers higher specificity as this dye is only liberated from the probe after binding to the fragment of interest. No fluorescence is created if no hybridisation occurs with the sequence of interest. Taqman® does not offer the ability of obtaining a melting curve and due to the high specificity is not suitable for detection of isomiRs, identical mature miRNAs that are generated from different genomic locations, whereas SYBR green can amplify isomiRs due to the lower specificity during reverse transcription and PCR stages. High sequence specificity of the Taqman® reverse transcription stem-loop primers and PCR probes, designed to bind to one unique precursor during reverse transcription and in the PCR phase to amplify one unique mature sequence, confers this specificity to Taqman®. As isomiRs are encoded in the sequence of different precursors in different parts of the genome, only the pair selected for Taqman® amplification would be quantified and therefore there is the disadvantage of underestimating the total amount of a mature miRNA sequence (including isomiRs). Fluorescence in both SYBR green and Taqman is proportional to the amount of sequence of interest found in the reaction.

Exploiting the knowledge that some microRNAs are highly conserved across species, three groups investigated microRNA expression in equine muscular disease, breeding and in equine cartilage/bone disease using qRT-PCR with human miRNA sequences (Barrey *et al.*, 2010; da Silveira *et al.*, 2012; Buechli *et al.*, 2013). Barrey *et al.* (2010) investigated the abundance of 10 of the most commonly dysregulated microRNAs in human muscular disease. The study assessed horses with and without two common muscular pathologies, recurrent

exertional rhabdomyolysis (RER) and polysaccharide storage myopathy (PSSM). From this comparison they found a correlation between increased levels of six microRNAs and the presence of PSSM (miR-1, miR-133, miR-23a, miR-195 and miR-339 ($p<0.05$); data was normalized to an internal control (miR-375)). In a separate study examining miRNAs in breeding performance, da Silveira *et al.* (2012) found differences in expression of miR-181a, miR-375, miR-523-5p in young mares *vs.* older mares. This study was performed using SYBR green technology with primers designed to human miRNA sequences and expression levels were normalized against small nucleolar RNAs (U1, U6, U43). The three differentially expressed microRNAs identified have been found to regulate mitogen-activated protein kinase (MAPK). A third study conducted on an *in vitro* model, assessed the expression of miR-140 in equine mesenchymal cells using Taqman® (Buechli *et al.*, 2013). Expression levels of this microRNAs were normalized using U6 and results showed that the expression of this microRNA is linked to chondrogenic differentiation, and that mRNAs encoding C-X-C motif chemokine ligand 12 (CXCL12) and a disintegrin and metalloproteinase with thrombospondin motifs 5 (ADAMTS-5) may be targeted by miR-140.

1.2.5.3. Experimental verification of equine miRNAs using next generation sequencing

While studying the transcriptome of semen from five stallions, Das *et al.* (2013) identified 82 equine miRNAs. This was an interesting finding as library preparation before sequencing was not specific for miRNAs. miRNAs provided by the authors are annotated as pri-miRNA so it is difficult to know whether they identified only the precursor sequences or the mature miRNAs. Kim *et al.* (2014) used RNA seq technology to identify differences in miRNA profiles in three types of equine tissues (colon, muscle and liver). In total, 684 miRNA candidates were suggested to be present in the three tissues, of which 292 had been annotated in miRBase and 392 are suggested to be novel miRNAs, not present in the miRBase. When comparing the sequences of these putative miRNAs to the sequences available in miRBase, the variation in nucleotide composition and miRNA length is significant. Additionally, no experimental validation of the RNA seq data was performed by Kim et al., (2014) suggesting the likelihood of false positives in their list of miRNAs. Of the 292 mature miRNA candidates with annotation, 42

were present in all three tissues, 42 were specific to muscle, 119 of colon and 41 to liver.

The most recent and detailed experimental study on equine miRNAs with the most robust analysis to date was released soon after Kim *et al.* (2014) published their findings (Desjardin *et al.*, 2014). In this study, miRNAs were sequenced from cartilage and bone samples obtained from 10 month-old foals with and without a common orthopaedic condition, named osteochondrosis (OC). Post-mortem specimens were analysed while raw and after biomechanical experiments. In total, the authors identified 670 miRNAs of which 371 can be found in miRBase and 299 are suggested to be novel miRNAs. A high percentage of miRNAs was found in both tissues (561 miRNAs, 85%). OC had an effect on the miRNA expression in cartilage (49 miRNAs) and bone (41 miRNAs). Confirmation of the differentially expressed miRNAs was achieved by quantifying three miRNAs in each tissue using qRT-PCR with Taqman® technology. Although this validation added robustness to the study, close analysis of the annotations published in the paper reveals that some of the annotations given to the mature miRNA candidates are not present in miRBase and that some miRNA candidates have mismatches in sequence and length when compared to their putative homologues in miRBase.

The aforementioned studies have contributed to the identification of miRNAs in the horse, and their expression in different tissues and in a small group of pathologies (OC, PSSM, RER) as well as in breeding performance. The study of miRNAs in other species has helped to answer scientific questions about disease mechanisms including tumorigenesis and this can be potentially utilised to improve diagnostic tools and treatment options (Iorio & Croce, 2012). In addition, characterising the profile of miRNAs in equine sarcoids and identifying their potential regulatory function will be of benefit in understanding unknown aspects of the pathogenesis and aetiology of equine sarcoids.

1.3. Aims and objectives of the project

As fibroblasts are the main cell type altered in ES, this project utilised a number of technologies to examine any changes in expression of equine miRNAs in this cell type, which may help understand the aetiology of ES and potentially be applied to aid prognosis and diagnosis. The main aims were to:

1. Describe the miRNA profile in equine normal fibroblasts (equine primary fibroblasts)

No previous studies have been performed describing the miRNA profile in equine fibroblasts.

2. Study the differential expression of miRNAs in equine sarcoids using an *in vitro* model.

Since the presence of the BPV oncoproteins cause alterations in equine fibroblasts promoting tumoural transformation, and knowing that other viral induced cancers have an abnormal microRNA expression profile, this study aimed to identify cellular microRNAs that may be altered in equine sarcoids. Comparison of miRNA profiles was carried out using three different cell samples: 1) normal equine fibroblasts (EqPalFs), 2) cells transformed with the BPV-1 genome, EqPalF-BPV-1 cells referred to as S6-2 cells, and 3) equine tumoural fibroblasts, named EqSO4bs (Yuan et al., 2008a).

Chapter 2. Material and methods

2.1. General laboratory practice

Hazardous chemicals were handled as described in the Control of Substances Hazardous to Health (COSHH) guidelines from the University of Glasgow. Laboratory coat and powder-free nitrile gloves were worn at all stages. Details for disposables, plastic ware, reagents, other chemicals, kits and electrical equipment, including manufacturers/providers, are listed in this chapter in Tables 2.1. – 2.10.

2.2. Cell culture

Cell manipulation was performed in a designated tissue culture laboratory under aseptic conditions. With the exception of thawing cells and cell count, all procedures were performed in a Class II microbiological safety cabinet, BioMat 2 (Medical Air Technologies Ltd, Manchester, U.K.). Prior to being introduced into the hood and/or the incubator, the surfaces and exterior of container and vials were treated with 70% molecular grade ethanol (VWR International Ltd., Lutterworth).

2.2.1. Cell culture media, buffers and reagents

Cell culture media was mixed and aliquoted within a Class II microbiological safety cabinet and stored at 4°C. When ingredients were to be used more than once, these were aliquoted and stored at either 4°C or -20°C following manufacturer's advice. Recipes for the different types of media and manufacturers are displayed in Table 2.1. Commercially prepared media and plastic ware are displayed in Table 2.2. and Table 2.3. respectively.

Table 2.1. Description of recipes of media used in cell culture

Name	Final concentration	Abbreviation	Ingredients	Manufacturer
Culture media	88.50%	DMEM	High Glucose Dulbecco's Modified Eagles medium with Pyruvate - DMEM-GlutaMAX™	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
	10%	FBS	Foetal Bovine Serum, qualified, heat inactivated, E.U.-approved, South America Origin	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
	1%	Pen-Strep	Penicillin-Streptomycin (10,000 U/mL)	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
	0.50%	Fungizone	Fungizone® Antimycotic	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
Thawing media	90%	DMEM	High Glucose Dulbecco's Modified Eagles medium with Pyruvate - DMEM-GlutaMAX™	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
	10%	FBS	Foetal Bovine Serum, qualified, heat inactivated, E.U.-approved, South America Origin	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
Freezing media	45%	Culture media	Culture media	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
	45%	FBS	Foetal Bovine Serum, qualified, heat inactivated, E.U.-approved, South America Origin	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
	10%	DMSO	Dimethyl sulfoxide	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
Phosphate - buffered saline (PBS)	140 mM	NaCl	Sodium chloride	Sigma-Aldrich®, Dorset, U.K.
	2.7 mM	KCl	Potassium chloride	Sigma-Aldrich®, Dorset, U.K.
	10 mM	Na_2HPO_4	Sodium phosphate dibasic	Sigma-Aldrich®, Dorset, U.K.
	1.8 mM	KH_2PO_4	Monopotassium phosphate	Sigma-Aldrich®, Dorset, U.K.

Table 2.2. List of commercially sourced buffers and reagents used in cell culture

Name	Abbreviation	Use	Manufacturer
Hanks Buffered Salt Solution	HBSS	Cell wash	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
0.05% Trypsin-EDTA (1x)	Trypsin	remove cells from monolayer	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
0.4% Trypan blue solution	Trypan Blue	assess cell viability and cell count	Sigma-Aldrich®, Dorset, U.K.

Table 2.3. Plastic ware utilised in cell culture

Name	Abbreviation	Manufacturer
Cellstar® cell culture flasks, PS, red filter cap (sizes T25, T75 and T175)	culture flask (T25, T75 or T175)	Greiner-bio one, Gloucestershire, U.K.
Cellstar® 15 ml PP tubes, graduated, conical bottom, blue screw cap, sterile	15 ml tubes	Greiner-bio one, Gloucestershire, U.K.
1.5ml polypropylene reaction tube with cap, natural, free of detectable DNase and RNase	1.5 ml tube	Greiner-bio one, Gloucestershire, U.K.
2 ml, 5 ml, 10 ml and 25 ml serological pipette, individually wrapped, paper/plastic packaging, sterile, free of detectable DNase and RNase	Serological pipette	Greiner-bio one, Gloucestershire, U.K.
20 µl Graduated Filter Tips (Sterile)	Tips	StarLab, Milton Keynes, U.K.
2 ml Corning® cryogenic vials, external thread	2 ml Cryovials	Sigma-Aldrich®, Dorset, U.K.
50ml Centrifuge tube, conical base, bagged, sterile	50ml centrifuge tube	Greiner-bio one, Gloucestershire, U.K.

2.2.2. Description of cell lines

All cell lines used in this study were developed at the University of Glasgow (Yuan *et al.*, 2008a) and all tissues utilised for the establishment of the cell lines were taken with informed owner consent. A summary of the features of the three cell lines is presented in Table 2.4.

Table 2.4. Description of cell lines

Name of the cell line	EqPalFs	S6-2s	EqSO4bs
Description	equine primary fibroblast	equine primary fibroblast *in vitro* transformed with BPV-1 genomes (from equine sarcoid S6)	Equine tumoural fibroblast from a natural tumour
BPV-1 episomal genome	absent	present	present
BPV-1 E5 variant	Not applicable	Sw I	Sw I
BPV-1 LCR variant	Not applicable	II	VII
Contact inhibition	normal	lost	lost
Life span (passage number)	13	21	24
Population doubling (hours)	57.8	24	26.7
Source	Cell bank, University of Glasgow (described in Yuan *et al.* (2008a)	Cell bank, University of Glasgow (described in Yuan et al. (2008a)	Cell bank, University of Glasgow (described in Yuan et al. (2008a)

EqPalFs are equine primary fibroblast cells derived from a foetal palate of an aborted equine foetus. EqPalFs do not contain BPV-1 episomal genomes and because they are primary cells, they have a limited life span of 13 passages (Yuan *et al.*, 2008a).

S6-2s are EqPalFs transformed with BPV-1 episomal genomes. The BPV-1 genome used to transform EqPalFs was derived from an equine sarcoid tumour (S6) by Rolling Circle Amplification followed by restriction digestion, purification, cloning and recircularization (Yuan *et al.*, 2007c). S6-2 cells contain two BPV-1 genomes per cell of variant Sw I for E5 and variant II for LCR (Yuan *et al.*, 2008a). S6-2 cells express early and late BPV-1 gene transcripts as detected by RT-PCR. These cells are morphologically transformed, invasive and display lack of contact inhibition (Yuan *et al.*, 2008a).

EqSO4b cells are immortal cells derived from a naturally occurring fibroblastic sarcoid tumour confirmed by histopathological analysis following haematoxylin and eosin staining (Yuan *et al.*, 2008a). They contain eight BPV-1 episomal genomes per cell of the E5 variant Sw I and LCR variant VII. EqSO4b cells express early and late BPV-1 gene transcripts as detected by RT-PCR and they lack contact inhibition (Yuan *et al.*, 2008a).

All cell lines were kept in liquid nitrogen at -80°C in freezing media. Details of the freezing media can be found on Table 2.1.

2.2.3. Cell thawing, cell count and growth

Aliquots of 1-2 x 10^6 cells/ml stored in 2 ml cryovials were brought up from the liquid nitrogen containers taking the appropriate safety precautions. Cells were thawed in a 37°C water bath W6 (Grant Instruments, Cambridge) for less than one minute and the cryovials were then wiped with 70% ethanol and introduced it in the microbiological safety cabinet. The content of each cryovial was resuspended in 10 ml of pre-warmed thawing media prepared as described in Table 2.1. The cells were then centrifuged for 5 minutes at 1200 g in Heraceus® Multifuge® 3 SR (Thermo-Scientific, Renfrew, U.K) and the supernatant was completely removed. The cell pellet was then resuspended in 1 ml of pre-warmed culture media made up as described in Table 2.1. An aliquot of 25 µl of

cells resuspended in culture media was mixed with 25 µl of 0.4% Trypan Blue solution (Trypan Blue) (Sigma-Aldrich®, Dorset, U.K.) resulting in a two-fold dilution. Cells were assessed under the microscope at 40x magnification for cell viability and cells were counted in the 16 squares of the four blocks of a haemocytometer to calculate the average live cells (without Trypan Blue uptake) taking into account the two-fold dilution:

LC = (live cells) transparent cells in 16sq, in the 4 blocks.

FD = Fold dilution (=2).

vol = volume in which cell pellet was resuspended (=1ml).

TC (total cells)/ml= LC x ¼ x 10^4 x FD

Cells were seeded in a T25 culture flask at a density of 4x10^4 /cc^2 with 10 ml of pre-warmed culture media and maintained in a 37°C humidified atmosphere with 5% CO_2 in the CO_2 Incubator HERAcell® 150 i / 240 i (Thermo-Scientific). Culture flasks were labelled with the name of the cell line, cell seeding volume, passage and date. When cultures became 80% confluent, cells were passaged or used for downstream work.

2.2.4. Cell passaging, mycoplasma testing and sampling for DNA and RNA extractions

To detach monolayer cells, firstly supernatant from flasks was aspirated and discarded, and cells were washed with Hanks Buffered Salt Solution (HBSS) (Invitrogen™-Life Technologies Ltd, Renfrew, U.K.). Cells were then exposed to 0.25% Trypsin (Invitrogen™-Life Technologies Ltd) for less than 1 minute in the incubators at 37°C, and then resuspended in pre-warmed culture media. After centrifugation at 1200 g for 5 minutes, cells were evaluated and counted with Trypan Blue as previously described in section 2.2.3. (Cell thawing, cell count and maintenance) and seeded in flasks with 10 ml culture media of for T25 flasks, 25 ml for T75 and 50 ml for T125 flasks. Cells were routinely sent to microbiology department to be tested for mycoplasma using MycoAlert™ Mycoplasma detection kit (Lonza Biologics plc, Cambridge, U.K.) and they were found to be negative. After reaching 80% confluence cells were:

- Frozen in freezing media (for details on the freezing media please refer to Table 2.1.) to maintain the cell bank.
- Resuspended in HBSS/ Phosphate - buffered saline (PBS) for DNA extraction and used immediately.
- Resuspended in HBSS for immediate RNA extraction using the column method.
- Resuspended in Trizol® Reagent (Invitrogen™, Life-Technologies Ltd), vortexed and either used immediately for RNA extraction with the phenol method or kept at -80°C until RNA extraction with the phenol method.

2.3. Molecular biology

Details of plastic ware, kits, chemicals and electric equipment utilised in this section are displayed in Tables 2.5. -2.9.

2.3.1. DNA extraction

DNA from cells in culture (80-90% confluence in T75 flasks, approximately $8x10^6$cells) was isolated using a column method QIAamp® DNA Mini and Blood Mini Handbook (Qiagen®, Manchester, U.K.) following the manufacture's spin protocol for DNA purification from blood/body fluids. In summary, cells in monolayer were trypsinised as described in section 2.2.4. (Cell passaging, mycoplasma testing and sampling for DNA and RNA extractions). Cells were washed with HBSS and spun for 5 minutes at 1200 g. The pellet was resuspended in 200 µl of PBS and cellular components, histones and other proteins associated with the DNA were digested using Proteinase K included in the kit (10 µl proteinase K per 100 µl of cells in suspension). After adding 1:1 volume of lysis buffer (AL buffer) and incubating at 56°C for 10 minutes in a water bath Sub 36 (Grant Instruments) and pulse centrifuging the tube, 100% molecular grade ethanol was added at a 1:2 ratio. The samples were collected in the column with the use of two buffers (AW1 and AW2), spinning the sample between buffers. The last buffer was added to collect the extracted DNA in a clean tube. After centrifuging and removing the buffer, the DNA was resuspended in 200µl of Tris EDTA buffer (TE buffer) (Microzone, West Sussex, U.K.).

2.3.2. RNA extraction

2.3.2.1. RNA extraction using column method

RNA from cells in culture was obtained using miRNeasy Mini Kit (Qiagen®) following the manufacturer's instructions from the section 'Purification of Total RNA, Including Small RNAs, from Animal Cells'. Typically, no more than 3.0 x 10^6 cells in suspension of HBSS, were centrifuged and the cell pellet,after supernatant removal,was resuspended in 700 µl of QIAzol Lysis Reagent included in the kit, and homogenized by vortexing for 1 minute to lyse the cells. Homogenate was incubated at room temperature for 5 minutes to promote dissociation of nucleoprotein complexes. After addition of 0.2 volumes of chloroform, the sample was mixed for 15 seconds and centrifuged for 15 minutes at 12000 g at 4°C. Supernatant containing RNA was transferred to a fresh tube and 1.5 volumes of 100% ethanol was added. The mix was pipetted through a column to collect the RNA in the membrane of the column and washed with two buffers, RWT and RPE and centrifuged at 8000 g. Composition of these buffers is undisclosed by the manufacturer but their purpose was to wash the RNA off (RWT) and to prevent ethanol carry-over (RPE). The dried RNA retained in the column was eluted with 30-50 µl of RNAse free water provided in the kit. Total RNA obtained with the column method was kept at -80°C until further use.

2.3.2.2. RNA extraction using phenol based method

Cells in monolayer were removed from the culture flasks using 7 ml of Trizol® reagent (Life Technologies) per T75 flask. Once resuspended in Trizol® reagent, the cells were placed in a 14 ml high speed tube (for details please seeTable 2.5.) and vortexed for 1 minute and either:

- Kept at -80°C until further use, in which case on the day of the RNA extraction they were fast-thawed in a 25°C bath and processed as described below.

- Used immediately for RNA extraction.

The first step after thawing or removal from the culture flasks was 5 minutes of incubation at room temperature to allow dissociation of nucleoprotein

complexes. Phenol: chloroform: isoamyl alcohol (25:24:1) (chloroform) (Sigma-Aldrich Company Ltd., Dorset, U.K.) was added at a ratio of 0.2 volumes (1.4 ml per 7 ml of Trizol®) and samples were then mixed vigorously for 15 seconds and left at room temperature for 3 minutes. After centrifugation at 12000 g for 15 minutes at 4°C, a three phase mixture was obtained. The upper layer, aqueous and transparent contained the RNA and it was approximately 40-50% of the volume. The aqueous phase was transferred to a clean RNAse free 14 ml high-speed tube avoiding contact with the interphase and 100% isopropyl alcohol (IPA) (Sigma-Aldrich Company Ltd., Dorset, U.K.) was added (3.5 ml per 7 ml of Trizol®) to precipitate the RNA. Samples were then incubated at room temperature for 10 minutes and centrifuged for 5 minutes at 12000 g at 4°C. The RNA precipitate, a gel-like structure visible to the naked eye, was then washed. Firstly, the IPA was removed and 75% ethanol was added (7 ml of ethanol per 7 ml of Trizol®). Subsequently the samples were vortexed briefly and centrifuged at 7500 g for 5 minutes at 4°C. The ethanol was then removed and the samples were dried at room temperature for a maximum of 10 minutes. RNA was resuspended in 40-50 µl of RNAse free water (Just Water, Microzone) and incubated at 55°C for 5 minutes to facilitate RNA solubilisation. Samples were then stored at -80°C for further downstream work or temporarily kept on ice during assessment of integrity and quantification.

2.3.3. Assessment of purity, integrity and quantity of nucleic acids

2.3.3.1. DNA purity and quantification

The purity and quantity of total genomic DNA were assessed using optical density measurements obtained with the NanoDrop® ND-1000 spectrophotometer (Thermo-Scientific) using the standard protocol. After selecting the settings for double stranded DNA measurement (DNA-50, upper limit of detection 3700 ng/µl) and calibrating the spectrophotometer with TE buffer (running a 'blank'), a volume of 1.5 µl of DNA in solution was loaded in the lower measuring pedestal. After closing the sampling arm (upper measuring pedestal) and prompting the software to perform the reading, the values of DNA concentration (measured at 260 nm and expressed in ng/µl), ratios of absorbance at 260/280

nm wavelength and 260/230 nm wavelength were obtained. Samples were considered acceptable for downstream work with a 260/280 nm ratio of approximate 1.8 (indicating good DNA purity), a ratio 260/230 higher than the 260/280 and a unique peak of absorbance around 260 nm.

2.3.3.2. Polymerase Chain Reaction (PCR) of equine Glyceraldehyde-3-Phosphate Dehydrogenase (equine GAPDH) and BPV-1 E5 gene

For identification of host genomic DNA and to confirm presence/absence of BPV-1 episomal genomes, an aliquot of 200 ng of genomic DNA obtained from cells in culture was amplified by PCR using primers for an equine housekeeping gene, GAPDH, and for BPV-1 E5 gene. All PCR reactions were carried out in 0.2 ml Illustra™puReTaq Ready-To-Go PCR Beads (VWR International Ltd.) containing a recombinant polymerase and dNTPs All reactions were performed in the thermocyclers GeneAmp® PCR System 2400 and 2700 (Applied Biosystems, Paisley, U.K.). Each experiment with each set of primers had at least one negative control (all the components and Just water (Microzone Ltd.) instead of template) the samples of interest, and one positive control (equine GAPDH positive sample and BPV-1 E5 positive sample).

Equine GAPDH PCR

Primer sequences for equine GAPDH were obtained from Lange *et al.* (2013). Forward primer (5' GAG CTG AAT GGG AAG CTC AC 3') and reverse primer (5' CTG AGG GCC TTTCTCCTTCT 3') were added to each 25 µl PCR reaction to a final concentration of 0.2 µM. Cycling conditions suggested by Lange *et al.* (2013) were modified as follows: 5 minutes of denaturation at 94°C, 30 cycles of 1 minute at 94°C, 1 minute at 57°C and 1 minute at 72°C, followed by 10 minutes at 72°C. Samples were held at 4°C until collected from the thermocycler. The expected amplified DNA fragment of 393 bp corresponds to a segment of equine GAPDH (NCBI annotation NM_001163856, starting at 715 bp and ending at 1107 bp).

BPV-1 E5 gene PCR

Primer sequences for BPV-1 E5 gene were obtained from Brandt *et al*. (2008b). Forward (5' CAC TAC CTC CTG GAA TGA ACA TTT CC 3') and reverse primer (5' CTA CCT TWG GTA TCA CAT CTG GTG G 3') were added to each 25 µl PCR reaction to a final concentration of 0.2 µM. Cycling conditions suggested by Brandt et al. (2008) were modified as follows: 5 minutes of denaturation at 94°C, 40 cycles of 30 seconds at 94°C, 45 seconds at 58°C, 45 seconds 72°C, followed by 5 minutes at 72°C. Samples were held at 4°C until collected from the thermocycler. The expected amplified DNA fragment of 499 bp corresponds to a segment of BPV-1 (NCBI annotation X02346.1, starting at 3788 bp and ending at 4286 bp).

2.3.3.3. Agarose gel electrophoresis of DNA fragments

In order to visualise and verify the size of the amplified DNA, the PCR products were subjected to agarose gel electrophoresis. Typically 0.5g of ultra-pure agarose was mixed with 50 ml of Tris acetate-EDTA buffer (TAE buffer, for details please see Table 2.7.) and the mix was warmed until the agarose was fully dissolved and the solution was transparent. After the temperature decreases below 50°C, 2 µl of Ethidium bromide (10mg/ml) was dissolved in the agarose-TAE buffer mixture. The solution was allowed to solidify in a casting tray for 30 minutes, with a gel comb (9 to 15 wells). Once the gel solidified, the PCR products were loaded into a well with loading dye (for details on the loading dye please see Table 2.7.). The products were run at 70-80 volts over 60 minutes on the Mini-Sub® Cell GT Cell (Bio-rad Laboratories, Hertfordshire, U.K.), an electrophoresis unit, and the gel was visualised on a UV transilluminator (UVi tec, Thistle Scientific Ltd., Glasgow, U.K.). The size of the fragments was assessed in comparison to the 100 bp DNA ladder (Invitrogen™).

2.3.3.4. Agarose gel purification and DNA sequencing analysis of the PCR products

PCR products were excised from the gel and purified using QIAquickPCR purification kit protocol (Qiagen®). This method uses a filter that captures the DNA within the column and washes the rest of the components using various

buffer solutions (named PB and PE). Instead of collecting the eluted DNA in EB buffer as per manufacture's advice, PCR products were eluted in 30-50 µl of Just water (Microzone Ltd.). After purification, fragments were assessed by gel electrophoresis again on a 1% agarose-TAE buffer gel and sent for DNA sequencing by a commercial company (Source Bioscience Sequencing , Belhill, U.K.). Each product was sequenced using either amplicon-specific forward or reverse primer. Sequences were searched against the NCBI database (Altschul *et al.*, 1990) using BLAST to verify the nucleotide sequence matched equine GAPDH (NM_001163856) or BPV-1 E5 (X02346.1)

2.3.3.5. RNA integrity using agarose gel electrophoresis

Integrity of total RNA was assessed using 1 % agarose gel electrophoresis in TAE buffer. Electrophoresis tank, tray and lid were treated with RNAse zap (Sigma-Aldrich®) and rinsed with RNase free water. The method of preparation of the gel and electrophoresis, except for the initial RNAse zap treatment, is identical to the protocol used for PCR product visualisation (section 2.3.3.1.2. - agarose gel electrophoresis).

RNA Integrity was assessed by comparing the ribosomal bands of the 28S and 18S to the bands from a 100 bp ladder (Invitrogen™-Life Technologies)

Intensity of the bands, comparison and size was used to assess RNA integrity and level of degradation. RNA was considered of good quality when the 18S ribosomal unit ran at 4.5 Kb, the 28S at 1.2 Kb and the intensity of 28S band was twice as the intensity of 18S. The mRNA will appear as a smear from 0.5–12 kb.

2.3.3.6. RNA quantitation using spectrophotometry (NanoDrop® ND-1000 Spectrophotometer)

Quantity screening of RNA was performed with the NanoDrop® ND-1000 Spectrophotometer (Thermo-Scientific, Pailsey, U.K.). This quantification method cannot distinguish between DNA and RNA so RNA values can be artificially overestimated when using this technique. However, this method of quantitation aids downstream work, such as DNase treatment, where the value of total nucleic acid concentration is needed.

In summary, samples (2 µl) were loaded on the spectrophotometer after calibration and concentration of nucleic acids was noted. Samples were considered acceptable when one unique peak was noted at 260 nm and when the 260/280 ratio was approximately 2.0. A 260/280 ratio of 2.0 was considered acceptable. Samples with lower 260/280 and/or 260/230 ratios were discarded due to possible protein, phenol or ethanol contaminants that may affect downstream work.

2.3.3.7. RNA quantitation using fluorometry (Qubit™ fluorometer)

Qubit™ fluorometric quantitation (Thermo-Scientific) is a method that uses a fluorescent agent that specifically and uniquely binds to RNA. The samples in this project were measured using Qubit™ RNA BR assay following the supplier's recommendations. Firstly a working solution was prepared that contains a buffer and a fluorescent reagent. Secondly the working solution was distributed in 0.5 ml Qubit tubes (for more details see Table 2.5.): the tubes labelled as standard (needed for calibration) contained 199 µl of working solution and 1 µl of standard whereas the tubes labelled for the RNA samples contained 198 µl of working solution and 2 µl of RNA sample. The tubes were then introduced into the Qubit unit and after calibration with the standards 1 and 2, fluorescence readings were obtained.

When samples exceeded the range of the RNA BR kit, samples were diluted 1:6 and the reading was repeated. RNA samples were stored at -20˚C for short term use or at -80˚C for long term use. RNA aliquots were diluted in RNAse free water to fit the concentration needed for further downstream work (SYBR and Taqman qRT-PCR).

Table 2.5. Disposables utilised during RNA extraction, DNA extraction, RNA/DNA quantitation/integrity assessment, cDNA first strand synthesis and qRT-PCR

Name of the item	Abbreviation	Manufacturer
14 ml polypropylene (PP) tubes, round bottom, two-position vent stopper, sterile, single packed	14 ml high speed tubes	Greiner-bio one, Gloucestershire, U.K.
1.5ml polypropylene reaction tube with cap, natural, free of detectable DNase and RNase	1.5 ml tube	Greiner-bio one, Gloucestershire, U.K.
PCR microtubes with flat cap, 0,5 ml	0.5 ml Qubit tubes	VWR International Ltd., Lutterworth
2 ml, 5 ml, 10 ml and 25 ml serological pipette, individually wrapped, paper/plastic packaging, sterile, free of detectable DNase and RNase	Serological pipette	Greiner-bio one, Gloucestershire, UK
10µl, 20µl, 200µL and 1000µL filter tip	filtered tips	StarLab, Milton Keynes, U.K.
0.2 ml Illustra™puReTaq Ready-To-Go PCR Beads	PCR beads	VWR International Ltd., Lutterworth
PCR plates - non skirted	Stratagen plates	StarLab, Milton Keynes, U.K.
MicroAmp® Optical 8-Cap Strip	Cap strips	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
MicroAmp® Fast Optical 96-Well Reaction Plate	7500 plates	Invitrogen™-Life Technologies Ltd,Paisley, U.K.

Table 2.6. Kits utilised during RNA extraction, DNA extraction, RNA/DNA quantitation/integrity assessment, cDNA first strand synthesis and qRT-PCR

Name	Method	Provider
QIAamp® DNA Mini and Blood Mini Handbook	DNA purification	Qiagen®, Manchester, U.K.
Tris EDTA buffer (TE buffer)	DNA purification and Nanodrop calibration	Microzone, West Sussex, U.K.
miRNeasy Mini Kit	miRNA purification	Qiagen®, Manchester, U.K.
QIAquickPCR purification kit protocol	PCR product purification	Qiagen®, Manchester, U.K.
RNA BR fluorometric quantitation (Qubit™)	Quantification of RNA	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
DNA-free™ kit	DNase treatment of RNA	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
Mir-X™ miRNA First-Strand Synthesis kit	SYBR green first strand synthesis	Clontech, Takara Bio Europe SAS, Saint-Germain-en-Laye, France
SYBR® qRT-PCR kit	SYBR green qRT-PC master mixes	Clontech, Takara Bio Europe SAS, Saint-Germain-en-Laye, France
TaqMan® MicroRNA Reverse Transcription Kit	Taqman RT master mix	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
TaqMan® MicroRNA Assays (Inventoried)	Taqman RT primers and qPCR primers and probes	Invitrogen™-Life Technologies Ltd,Paisley, U.K.

Table 2.7. Buffers, markers and carriers utilised in gel electrophoresis

Item	Recipe or manufacturer
1kbp ladder	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
100 bp ladder	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
Loading dye	50% glycerol, 0.25% bromophenol blue, and 0.25% xylene cyanole in 1x TE buffer
Tris-acetate buffer (TAE) buffer	Tris base 484.5g, NaOAc 272.15 g, NaCl 116.8 g, Na2EDTA 74.45 g. pH adjusted to 8.15 with glacial acetic acid and made up to 2 L volume.

Table 2.8. Other chemicals utilised during RNA extraction, DNA extraction, RNA/DNA quantitation/integrity assessment, cDNA first strand synthesis and qRT-PCR

Name	Abbreviation	Method	Provider
Trizol®	Trizol®	total RNA purification	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
Phenol: chloroform: isoamyl alcohol (25:24:1)	Chloroform	total RNA purification	Sigma-Aldrich Company Ltd. (Dorset, United Kingdom)
Isopropyl alcohol (Propan-2-ol or Isopropanol)	IPA	total RNA purification	Sigma-Aldrich Company Ltd., Dorset, U.K.
Ethanol absolute AnalaR NORMAPUR® ACS, Reag. Ph. Eur. analytical reagent	Ethanol	DNA and RNA purifications	Sigma-Aldrich Company Ltd., Dorset, U.K.
Just water - 1ml sterile vials	Just water	total RNA purification	Microzone Ltd., West Sussex, U.K.
RNAse- zap	RNA-zap	RNA handling	Sigma-Aldrich Company Ltd., Dorset, U.K.
Ultra-Pure agarose	agarose	Gel electrophoresis	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
Ethidium Bromide	EB	Gel electrophoresis	Invitrogen™-Life Technologies Ltd,Paisley, U.K.

Table 2.9. Electrical equipment

Instrument	Abbreviation	Manufacturer
Class II microbiological safety cabinet, BioMat 2	Laminar flow hood	Medical Air Technologies Ltd, Manchester, U.K.
CO_2 Incubator HERAcell® 150 i / 240 i	Incubator	Thermo-Scientific, Renfrew, U.K.
water bath W6	Bath - cell culture	Grant Instruments, Cambridge, U.K.
Heraceus® Multifuge® 3 SR	Centrifuge cell culture	Thermo-Scientific, Renfrew, U.K
water bath Sub 36	Bath - Nucleic acids extraction	Grant Instruments, Cambridge, U.K.
CPR Centrifuge (Beckman)	Centrifuge - DNA extraction	Beckman Coulter, High Wycombe, U.K.
Avanti® J-E High-Performance Centrifuge	Centrifuge - RNA extraction	Beckman Coulter, High Wycombe, U.K.
NanoDrop® ND-1000 spectrophotometer	Nanodrop	Thermo-Scientific
GeneAmp® PCR System 2400	Thermocycler	Applied Biosystems, Paisley, U.K.
GeneAmp® PCR System 2700	Thermocycler	Applied Biosystems, Paisley, U.K.
Mini-Sub® Cell GT Cell	Electrophoresis tank	Bio-rad Laboratories, Hertfordshire, U.K.
Qubit™ fluorometer	Qubit	Invitrogen™-Life Technologies Ltd,Paisley, U.K.
Bioanalyzer 2100	Bioanalyzer 2100	Agilent Technologies UK Ltd, Edinburgh, U.K.
Mx3000PTM Real-Time PCR System (Stratagene)	Stratagene	Agilent Technologies UK Ltd, Edinburgh, U.K.
Applied Biosystems Fast 7500	Fast 7500	Applied Biosystems, Paisley, U.K.
UV transilluminator	UV box	UVi tec, Thistle Scientific Ltd., Glasgow, U.K.

2.3.3.8. RNA quantitation, integrity and sizing using automated electrophoresis and flow cytometry (Bioanalyzer 2100 (Agilent Technologies))

Total RNA samples used for microarray and for RNA-seq were assessed using 2100 Bioanalyzer at University of Glasgow Polyomics Facility. The chip was loaded with 1µl total RNA and size, quantity and integrity were examined. The bioanalyzer determines concentration of RNA and, based on the number of peaks and size of the peaks as well as the graph pattern, the Bioanalyzer derives an integrity value named RIN. RNA samples with RIN values of less 8.5 (out of 10) were discarded due to the risk of poor quality RNA and/or degradation. Total RNA with small or undetectable peaks corresponding to the small RNA fraction (that includes microRNAs) were also discarded.

2.4. Microarray

When cells grew to 80-90% confluency on a T175 flask, one total RNA sample from EqPalF cells and one sample from S6-2 cells was extracted using the phenol method. After samples being assessed for integrity, purity and quantity using 2% agarose gel electrophoresis and NanoDrop® ND-1000 Spectrophotometer as described in section 2.3.3.2.2., RNA samples were shipped on dry ice to an external company, Exiqon A/S (Vedbaek, Denmark) for microarray analyses. Upon arrival, samples were assessed again for integrity and quantity using a Bioanalyzer 2100. Total RNA with RIN values of 10.0 were utilised for the microRNA microarray. Three samples from the same RNA extraction were aliquoted and used as replicates for the microarray. An aliquot of 750 ng total RNA from both sample and reference were labelled with Hy3™ and Hy5™ fluorescent label, respectively, using the miRCURY LNA™ microRNA Hi-Power Labelling Kit, Hy3™/Hy5™ (Exiqon, Denmark) following the procedure described by the manufacturer. The reference sample used comprised a pooled sample of total RNA from both cell lines. Each well of the microarray was loaded simultaneously with sample and reference. The Hy3™-labelled samples and a Hy5™-labelled reference RNA sample were mixed and hybridized for 16 hours to the miRCURY LNA™ microRNA Array 7th Gen (Exiqon), using 2164 miRNA probes from human, equine (predicted miRNAs), viral and synthetic sequences from miRBase 19.0 (Griffiths-Jones *et al.*, 2006) including repeated probes for quality control. Image acquisition was performed using the signal intensity of the dyes and the slides were scanned using the Agilent G2565BA Microarray Scanner System (Agilent Technologies, Inc., USA). The image analysis was carried out using ImaGene® 9 (miRCURY LNA™ microRNA Array Analysis Software, Exiqon, Denmark). Pre-processing of the data included background subtraction (Normexp with offset value 10) and normalisation using the global lowess regression algorithm. Moderated t-statistics were applied to determine differential expression of microRNAs between EqPalF and S6-2 cells. Following this, the Benjamini and Hochberg multiple testing adjustment method was applied to the *P* values (Benjamini & Hochberg, 1995). *P* values of 0.05 or lower were considered to be statistically significant. Calculations were performed with R/Bioconductor using the limma package (Smyth, 2004). All the data analyses

were performed by Exiqon and supplied as an xlx file. provided in the DVD contained in this thesis.

2.5. miRNA sequencing using high throughput sequencing (HTS)

Total RNA samples from EqPalF cells, EqSO4b cells and S6-2 cells extracted using the phenol method (section 2.3.2.2.) were assessed for integrity, purity and quantity using Bioanalyzer 2100. As per Illumina advice, total RNA with RIN values of >8.0 were utilised for microRNA sequencing. High-throughput sequencing (HTS) was performed on three technical replicates from the same RNA preparation for EqPalF cells, and on three biological replicates from three different RNA preparations for EqSO4b and S6-2 cells (in total three libraries per cell line).

2.5.1. Library construction, cluster generation and sequencing

Library construction was carried out using 1-2 μg of total RNA and TruSeq® Small RNA library kit (Illumina Inc, San Diego, CA, USA) following the manufacturer's advice. Firstly the 3'-end adapter was added to each sample using T4 mutant ligase contained in the kit. Secondly, the 5'-end adapter was ligated in a similar manner. The 3'-end adapter targets the hydroxyl group exposed after Dicer cleavage and it is the DNA binding site for the RT primers. The first strand of cDNA is synthesised during a reverse transcriptase reaction and it was utilised as a template for PCR amplification using PCR primers with indexes (tags) to sequence multiple samples in the same run (multiplexing). At this stage a check point is recommended by the manufacturer and the PCR products were run on the high sensitivity DNA chip to ensure the samples contained suitable quality and proportion of miRNAs. After amplification, the library underwent size selection followed by polyacrylamide gel electrophoresis (PAGE) purification. Fragments in the range 145-160 nucleotides, as identified by comparison to the ladder, were selected. This band contains mature microRNA sequences up to 30 nucleotides together with the adaptors. After purification, microRNAs were precipitated with ethanol. Before sequencing, samples were run a second time

on a high sensitivity DNA chip on the Bioanalyzer to ensure a sufficient small RNA fraction was contained in the samples (band between 110 bp and 230 bp).

Cluster generation on solid phase was performed on the libraries (Metzker, 2010) followed by massive parallel sequencing in the Genomic Analyzer IIx (Illumina®) (for EqPalF) and NextSeq500 (Illumina®) (for EqSO4b and S6-2). The samples were run as a multiplex on a single-index single read sequencing run as recommended in the protocol.

2.5.2. Data analysis

An overview of the steps of the data analysis is displayed in Figure 2.1. Summary of the pipeline for analysis of the HTS data.

2.5.2.1. miRDeep2

Raw data was obtained in .fastq format and the files were pre-processed before being analysed with miRDeep2, a tool for miRNA identification (Friedländer *et al.*, 2012). The pre-processing of the files included transforming the sequencing files to .fasta format, removing the adapter sequences using Clip, generating a version of the equine genome (*Equus caballus*, EqCab2.0) with the appropriate format for miRDeep2 – downloaded from Ensembl, and formatted using BowTie. miRDeep2 software contains seven scripts and it is designed to process data from libraries removing redundant sequences (sequences that match fragments of SNORDs, tRNA and other non-small RNAs), removing small sequences (<18 nucleotides long), removing ubiquitous sequences (present in more than 5 genomic locations), aligning the results against the equine genome and equine miRNAs from the miRBase 20.0, predicting new miRNAs based on the folding energy (using RandFold) of the precursor sequence. The output from miRDeep is a .html file with a list of potential mature microRNA candidates together with their genomic location, mature and precursor sequences as well as read counts from the sequencer and miRDeep scores. Additional files contain graphical representations of the microRNAs with the predicted folding pattern in .pdf format.

Post-processing of the data obtained from miRDeep2 included harmonizing the data to generate a list of unique potential mature microRNA candidates from the three cell lines using a script and commands on command line. The following criteria were applied to generate a file of microRNAs:

1. If two mature sequences were encoded on the same chromosomal location with a maximal deviation of location of ten nucleotides, they are defined as the same microRNA (tolerance for chromosomal location 10).
2. If two mature sequences were the same length and had the same nucleotide sequence they were considered the same microRNA (i.e. tolerance for sequence composition 0).

The genomic location condition was applied first and nucleotide sequence followed. This was to prevent discarding the historically named 'star miRNA' (miRNA* or passenger strand), as miRDeep2 gives priority to the guide strand (for more details on miRNA biogenesis and miRNA *vs* miRNA*, the reader is referred to section 1.2.2.). Each potential microRNA candidates was assigned an internal ID.

2.5.2.2. Discovery of novel miRNAs

The list of internal IDs and mature sequences were further analysed using two methods to assign a name for the microRNA and if no known annotation was found, it was attempted to group them into a known family based on the sequence of the seed region. An overview of the steps is displayed in Figure 2.2.

Method 1 – BLAST

The list of miRNA candidates and their mature harmonized sequences was formatted to .fasta format using TextWrangler and compared against five databases using BLAST + (Camacho *et al.*, 2009) in Unix environment using Mac Command Line. The entire list of mature microRNAs downloaded from miRBase 20.0 (mature), the entire list of mature equine microRNAs (downloaded from the miRBase 20.0 and using command line to extract the equine entries)(eca), and the 3 previously published lists of microRNAs in the horse (Zhou *et al.*, 2009; Desjardin *et al.*, 2014; Kim *et al.*, 2014) were the five databases used for this

method. The list of potential annotations for each miRNA obtained from BLAST+ was processed using a script to select the top hits using maximal identity. Matches that had the following features were selected as 'perfect match': 100 % sequence identity, 100 % coverage with allowance of two mismatches (length of the query miRNA and length of the suggested annotation is either identical or up to two nucleotides of difference) and no internal mismatches in nucleotide sequence. This selection was performed on Excel.

Method 2 – Edit distance

The list of miRNA candidates and their mature harmonized sequences was compared to the same databases (mature, eca, Zhou, Kim and Desjardin) using a script and incorporating both nucleotide sequence and length as criteria for the comparison. The output generated consists of 'edit distances' – a number that reflects the number of mismatches either in length or in nucleotide composition - and a list of potential matches. Matches that had 2 or fewer edit distances were considered a 'perfect match'.

Seed region study

microRNA candidates were compared to previously described miRNAs using the seed region in an attempt to group them into previously described families. This was important for potential novel microRNAs to ensure no previously annotated name was found. This was achieved by using the list of microRNA families in miRBase and TargetScan. Harmonised data obtained from both databases was utilised to create a family database (seed region database) containing: the name of the family and the annotation (name and number in the form of MIPF0000002), precursor annotation (name and number in the form of MI0000001), mature 5' and mature 3' annotations (name and number in the form of MIMAT0000001) and seed regions for each entry (nucleotide positions 2-8). Seed regions of the miRNAs that had no perfect match in any of the databases after using method 1 and method 2, were compared to the seed region database to group unidentified miRNAs into previously described families.

2.5.2.3. Descriptive analysis

Correlation across replicates was calculated using Pearson's coefficient. Calculations were performed using Microsoft Excel. Percentage of GC nucleotides was calculated using a matrix on Microsoft Excel and miRNA length and CG percentage (%) were described.

Distribution of miRNAs across the genome was also studied and compared to data available from the horse (Zhou *et al.*, 2009; Desjardin *et al.*, 2014; Kim *et al.*, 2014).

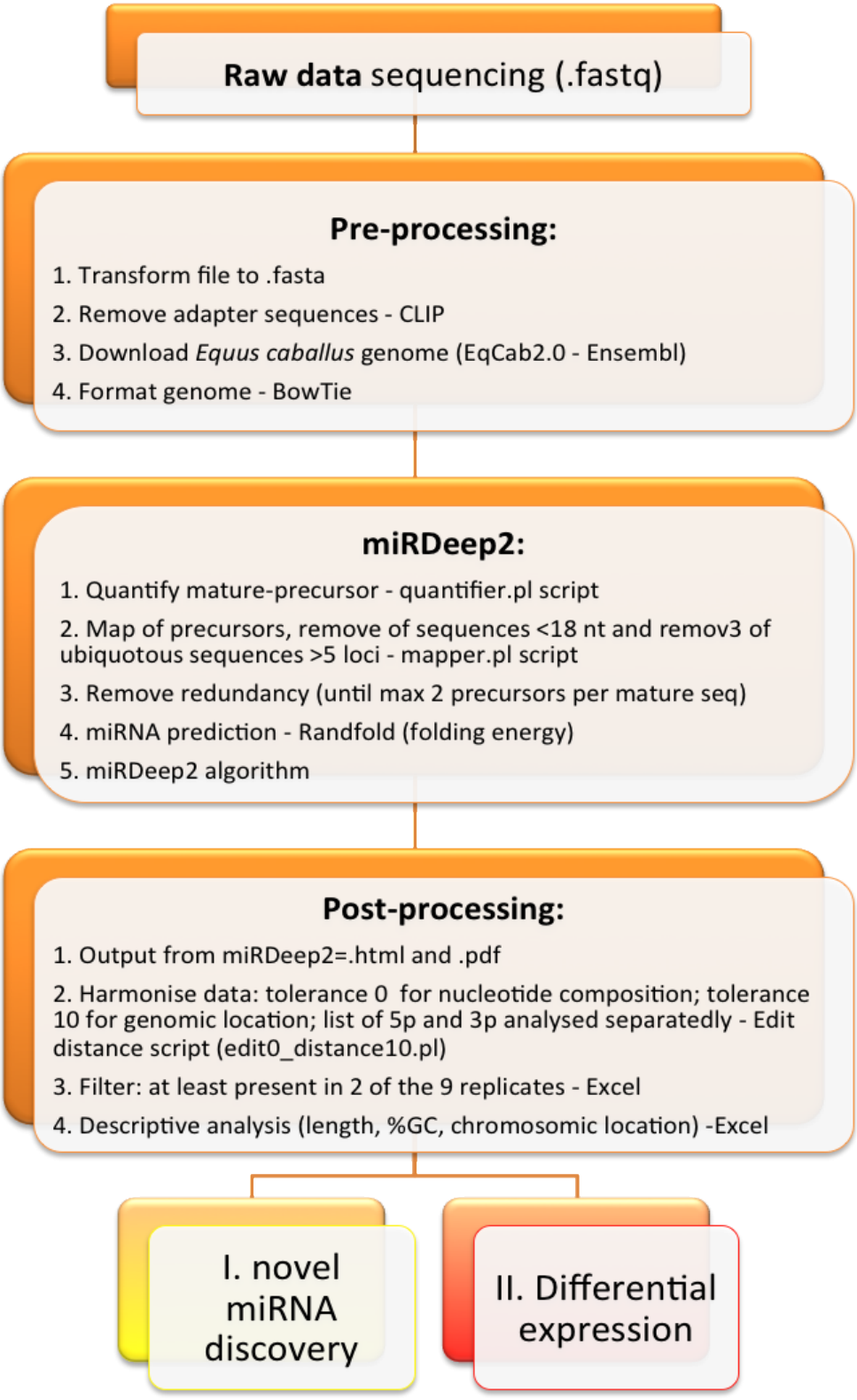


Figure 2.1. Summary of the pipeline for analysis of the HTS data

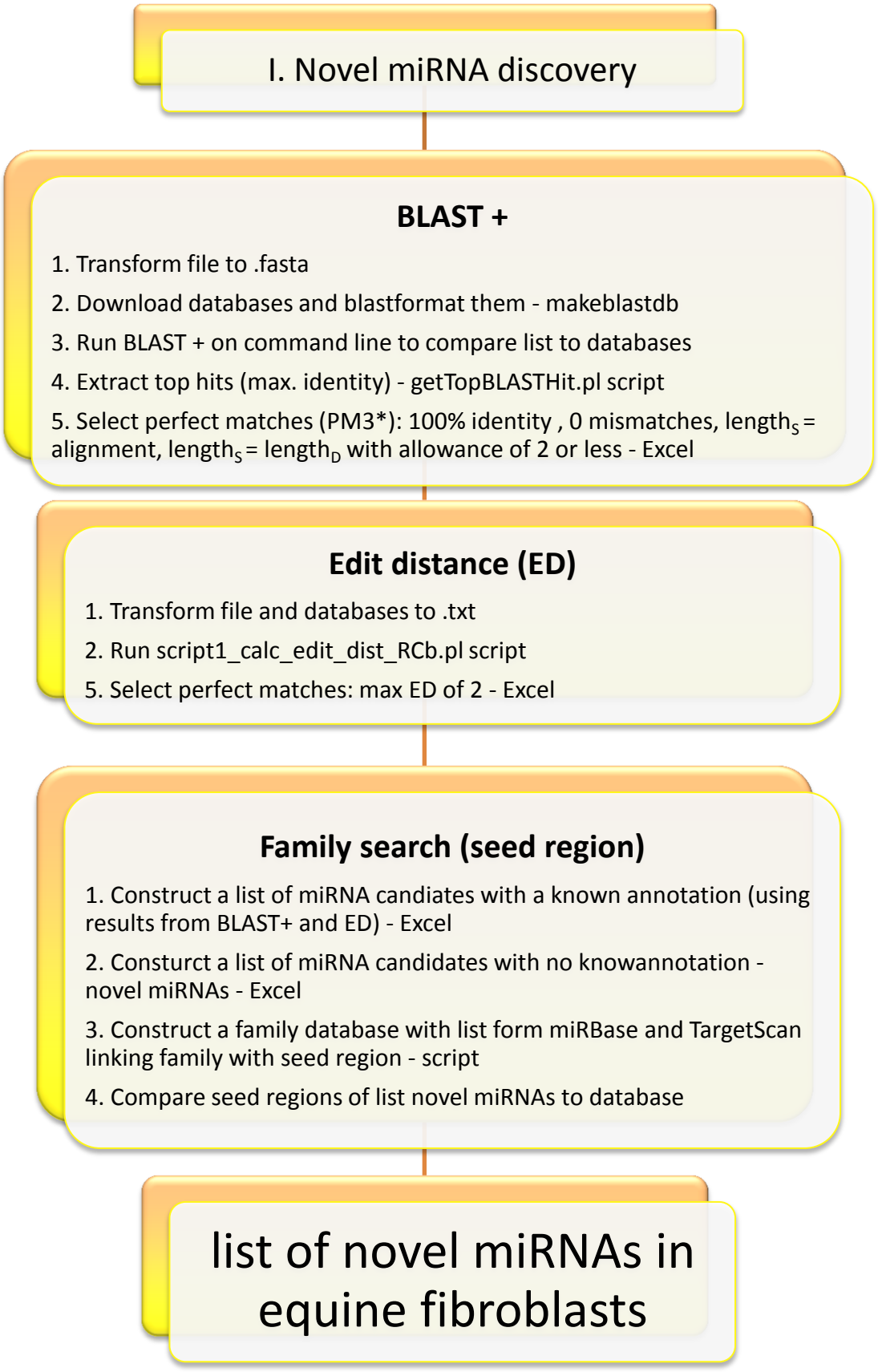


Figure 2.2. Summary of the pipeline for discovery of novel miRNAs

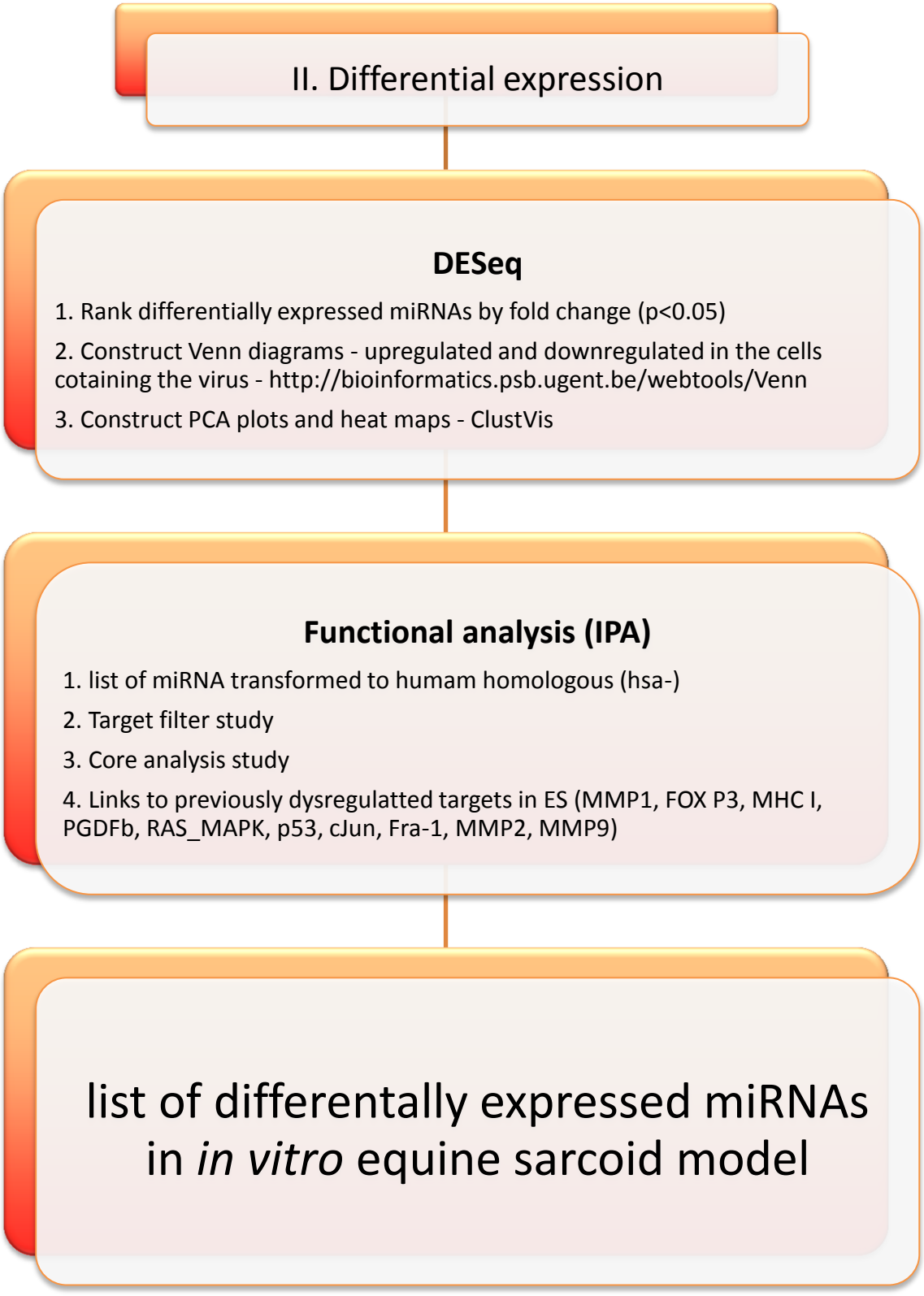


Figure 2.3. Summary of the steps of the analysis of differential expression of miRNAs in an equine sarcoid *in vitro* model

2.5.2.4. Differential expression

The number of reads from the cell lines was normalised to 20 million and differential expression of miRNAs was assessed using DESeq software. Significance was set at padj<0.05 for each pair compared (EqSO4b *vs.* EqPalF and S6-2 *vs.* EqPalF).

Differences were studied using PCA plot, heat maps and Venn diagrams. Functional analysis was performed using Ingenuity Pathway Analysis software (IPA) using core analysis, whereby pathways and networks altered as a result of dysregulated miRNAs, were highlighted.

2.6. Real-time, reverse transcription quantitative PCR (real-time RT-qPCR)

2.6.1. DNase treatment

Total RNA to be used for real time PCR analysis was treated with DNase I using a DNA-free™ kit (Invitrogen™-Life Technologies Ltd) following manufacturer's instructions. DNase treatment is recommended to eliminate any traces of genomic DNA that may overestimate the level of miRNAs in the fraction of expressed miRNAs.

Readings from the Nanodrop were used to dilute total RNA to a maximum of 200 µg/µl in RNAse free water, before proceeding with the 'Routine DNase treatment' protocol. Diluted RNA was transferred to nuclease free 0.5 ml tubes and 0.1 volumes of DNase buffer was added together with DNase (2 units per 10 µg of RNA per 50 µl reaction) and samples were incubated at 37°C for 30 minutes in the heat block. Deactivation of the DNase was achieved by adding 0.1 volumes of DNase inactivation reagent. After mixing well and incubating for 2 minutes at room temperature, samples were centrifuged at 10000 g for 1.5 minutes at 4°C. This step pellets the DNase and the DNase inactivating reagent and the upper layer contains RNA. RNA in suspension was transferred to a fresh 1.5 ml eppendorf without disturbing the interphase or the pellet, to prevent carryover of DNase inactivation reagent or DNase. Integrity of DNase treated

total RNA was assessed on a 1% agarose gel as described above (section 2.2.3.1.).

2.6.2. Relative quantitation of miRNAs with real-time qRT-PCR using SYBR green as a DNA dye binder (one step RT, RNA as template)

Using a Mir-X™ miRNA First-Strand Synthesis kit, 300 ng of total RNA from each cell line, were reverse transcribed to cDNA following the manufacturer's advice. In summary, during one step reaction, total RNA was polyadenylated and reverse transcribed. The final volume of resuspension was the only change implemented to the original protocol (50µl rather than 90 µl). Samples were aliquoted and kept at -20°C until further use. Using SYBR® qRT-PCR kit (Clontech, Takara Bio Europe SAS, Saint-Germain-en-Laye, France), each miRNA was amplified using a 3' end universal primer from the kit and a 5'-end specific primer designed following the advice from the manufacturer (i.e. entire sequence of the mature microRNA). Details of the primers are given in Chapter 3. The Master mix for each set of two primers, including the normalizer gene RNU6, was prepared for 28 reactions to account for pipetting error. Changes implemented to the protocol were volume of water (from 9 µl to 10.36 µl), volume of 5' and 3'-end primers (from 0.5 µl to 0.32µl) and volume of template (from 2 µl to 1 µl). These changes were introduced during optimization. Each set of primers was tested in each cell line in triplicate, together with two repeats of the non-reverse transcription (-RT) control and 5 repeats of the no-template control (NTC). Each plate contained RNU6 as a normalizer gene. Plates were loaded in Mx3000PTM Real-Time PCR System (Stratagene). Cycling conditions included a denaturation step of 10 seconds at 95ºC followed by 40 cycles at 95ºC for 5 seconds and 60ºC for 20 seconds. A dissociation curve was set at the end of the PCR for detection of non-specific amplification and/or primer dimer formation. Conditions of the dissociation curve were 60 seconds at 95ºC, 30 seconds at 55ºC and 30 seconds at 90ºC. The threshold cycle (Ct) or cycle number at which the fluorescence generated within a reaction crosses the threshold line, from each miRNA from each cell line, were normalised to an endogenous control. Ct values together with melting temperature of the PCR product were analysed and results of the

normalised data were expressed as a relative quantitation using the delta-delta C_t ($\Delta\Delta C_t$) method (see 2.6.4)

2.6.3. Relative quantitation with real-time qRT-PCR using probe-based technology Taqman® (two step RT, RNA as template)

DNase treated total RNA obtained using the phenol method was reverse transcribed and quantified using specific kits for microRNA from Life Technologies™ - Thermofisher (Taqman® microRNA assays). Each kit contains a specific RT stem loop primer for the microRNA of interest together with qPCR primers and probes, each of them specific for the mature microRNA of interest. In brief, 10 ng of total RNA was reverse transcribed to cDNA and then PCR amplified and quantified in real time using Applied Biosystems Fast 7500 real-time PCR system following manufacture's instruction. The initial amount of cDNA per PCR reaction was increased from 1.33 µl to 2 µl. PCR conditions were set at 15 seconds at 90ºC, followed by extension and annealing during 40 cycles at 60ºC for 60 seconds. Each amplification was performed in triplicate and the average Ct value was normalized using RNU6 (for more details please refer to chapter 4 and 5). Expression of each microRNA was quantified against the control cell line (EqPalF) using the ΔΔCt method.

2.6.4. Delta-delta C_t method ($\Delta\Delta C_t$ method).

The relative levels of expressed microRNAs obtained using SYBR green or Taqman technology were compared using the $\Delta\Delta C_t$ method (Winer *et al.*, 1999a). Normalised Ct values (using non-coding small nucleolar RNA component of U6 or RNU6) were compared using EqPalF as a calibrator (or reference sample) assuming primer efficiencies of 2 as described in the literature (Winer *et al.*, 1999a). The result was a ratio between the cell line of interest and the calibrator per each microRNA. Results were calculated using this formula:

Relative quantitation of a specific microRNA= $2^{-\Delta C_t (\text{EqSO4b or S6-2})} / 2^{-\Delta C_t (\text{EqPalF})} = x$

Where (ΔC_t (EqSO4b or S6-2)) is the normalized C_t of a specific microRNA in one of the two cell lines (EqSO4b or S6-2 cells), obtained with the subtraction of the C_t of the microRNA minus the C_t of the normalizer for that cell line; (ΔC_t

(EqPalF)) is the normalized C_t of that same microRNA in the primary cell line (named calibrator) and x is the result of this ratio.

Chapter 3. Differential expression of microRNAs in bovine papillomavirus type 1 transformed equine cells using microarray profiling

3.1. Introduction

The mechanisms of tumoural transformation by papillomaviruses are of significant interest and have been well studied, particularly in human medicine. To date, 181 types of human papillomaviruses (HPV) have been described and thirteen of these, referred to as high risk HPVs (HR-HPVs) by the World Health Organisation (WHO), carry a risk of inducing malignancies (WHO, 2012; Van Doorslaer *et al.*, 2013). Of particular importance are HPV -16 and HPV – 18, which belong to the group of Alpha HPVs, and are responsible for 4% of all cancers worldwide and 70% of cervical cancer and pre-cancerous lesions (Braaten & Laufer, 2008; Stanley, 2010; de Martel *et al.*, 2012). Based on evidence that supports their aetiologic association with human cancer, they are characterised as carcinogenic to humans in the latest report of biological agents by WHO (WHO, 2012). HPVs induce tumoural transformation by altering signalling pathways of two main tumour suppressors, p53 and pRB (Moody & Laimins, 2010; Doorbar *et al.*, 2015). p53 is a tumour suppressor that encodes a DNA binding protein that acts to decrease cell division. HPV E6 protein interferes with p53 function by promoting p53 degradation (through forming complexes of HPV E6 and cellular E3 ubiquitin ligase E6 associated protein (E6AP)) and by increasing p53 instability by inactivating p300 and Crebb-binding protein (CBP) (Münger *et al.*, 2004; Moody & Laimins, 2010). pRB is a member of the retinoblastoma family of proteins that control the G1-S phase transition by binding and decreasing the activity of E2F factors. E2F is a family of transcription factors involved in cell cycle regulation (cell cycle progression, differentiation, mitosis and apoptosis) and DNA synthesis (Moody & Laimins, 2010). Other factors contributing to the carcinogenic effect of HPVs are the alteration of telomerase, alteration of PDZ proteins and p21 and p27 and PI3K-AKT pathways (Moody & Laimins, 2010; DiMaio & Petti, 2013; de Freitas *et al.*, 2014).

It has been shown that these pathways are regulated by microRNAs (miRNAs), which are small non-coding RNAs and essential regulators at a cellular level (Gómez-Gómez *et al.*, 2013; Honegger *et al.*, 2015). The aberrant expression of miRNAs found in HR-HPVs-infected cells and cervical cancer samples has contributed to the understanding of the mechanisms by which the virus initiates

and perpetuates malignant transformation. In fact, gene expression profiling of miRNAs has identified aberrant expression of several miRNAs in HR-HPV-induced cancer in both *in vitro* and *in vivo* studies of cervical cancer and head-neck cancer (Wang *et al.*, 2008; Pereira *et al.*, 2010; Li *et al.*, 2011; Wald *et al.*, 2011; Lajer *et al.*, 2012; Wang *et al.*, 2014b). Dysregulation of miRNAs has been attributed to the presence of the HPV oncoproteins E6 and E7 (Lajer *et al.*, 2012; Ben *et al.*, 2015; Honegger *et al.*, 2015).

Several reviews have been recently published summarising the changes in miRNA expression in HPV induced cancer from microarray, qRT-PCR, Western Blotting and HTS data (Zheng & Wang, 2011; Gómez-Gómez *et al.*, 2013; Banno *et al.*, 2014; Sharma *et al.*, 2014; Wang *et al.*, 2016). Surprisingly, the profile of miRNAs associated with HPV infection has shown different results across studies. There is also variation between *in vitro* and *ex vivo* studies which is a reflection of the differences between the changes that HPV induces in cell culture *vs.* live tumours. A summary of the results from *in vitro* and *ex vivo* studies is displayed in Table 3.1. and Table 3.2. To add more complexity, different stages of HPV infection have been shown to change levels of miRNAs, verified by Wang *et al.* (2014b) using qRT-PCR. Furthermore, different stages of cervical cancer have an effect on miRNA profiles and levels, verified by Li *et al.* (2011) using microarray and by Gocze *et al.* (2013) using qRT-PCR technologies. Pereira *et al.* (2010) also reported that the level of variation in normal cervical samples was a major challenge to generating a profile specific for HPV induced cervical cancer.

Table 3.1. Levels of miRNAs found in 4 *in vitro* studies comparing HPV-16/-18 positive cell lines to control cells using microarray. miRNAs that show differences across studies are highlighted in bold in grey cells. The two rightmost rows show the results from *ex vivo* studies from HPV-induced cancer using microarray and the number of studies where each miRNA was reported. For details on the references from the *ex vivo* studies, the reader is referred to Table 3.2.

miRNA	Wang *et al.* (2008)	Wald *et al.* (2011)	Lajer *et al.* (2012)	Wang *et al.* (2014b)	Result of *in vivo* cervical cancer tissues	Number of studies
let-7a	Low				high	2
let-7b	Low				inconsistent	2
let-7c	High				high	2
let-7d	Low				high	2
let-7e	Low		low		high	1
let-7f	High		high		high	2
let-7g	**Low**		**high**		high	1
miR-100	**High**			**low**	low	4
miR-101	High				low	1
miR-106b				high	high	1
miR-10b	High				low	2
miR-1180			low		not reported	0
miR-125a-3p			low		not reported	0
miR-125b	High		high		low	3
miR-1275			low		not reported	0
miR-13232		low			not reported	0
miR-142-5p		low			high	1
miR-143	High				low	3
miR-145	High				low	4
miR-146a	High		high		inconsistent	2
miR-146b-5p	High		high		inconsistent	2
miR-150			high		not reported	0
miR-155		**low**	**high**		high	3
miR-15a	**Low**		**high**		high	2
miR-15b			high		high	2
miR-16			high	high	high	3
miR-181a	**High**	**low**			low	1
miR-181b	**High**	**low**	**low**		low	1
miR-181d	High				low	1
miR-191*	Low				inconsistent	1
miR-193b*			low		not reported	0
miR-195	Low				inconsistent	4
miR-198	low				high	1
miR-199a	high				low	3
miR-200b	low				high	1
miR-200c	low		low		high	1
miR-203	low				inconsistent	2
miR-205	high				high	3
miR-20b			high		high	2
miR-21	high		high		high	3
miR-210				high	not reported	0
miR-214	high				low	1
miR-218	**high**	**low**			low	3
miR-22	**high**			**low**	low	1
miR-221	**high**	**low**			inconsistent	2
miR-222	**high**	**low**			inconsistent	2
miR-223	low				high	2
miR-224				high	high	2

miR-23a	high				high	2
miR-23b	high				high	2
miR-24	**high**			**low**	low	1
miR-25				high	high	1
miR-26a	high				low	3
miR-26b	**low**		**high**		high	1
miR-27a	low			low	high	2
miR-27b	low				high	1
miR-29a	**high**	**low**	**high**	**low**	low	4
miR-301	high				low	1
miR-30a-5p	high				low	1
miR-30d	high				low	1
miR-30e-5p	high				low	1
miR-31			low		high	1
miR-31*			low		high	1
miR-324-5p			low		low	1
miR-33		high			not reported	0
miR-342-3p			high		not reported	0
miR-34a	high				low	1
miR-34a*			high		low	1
miR-34c	high				low	1
miR-363		high	high		not reported	0
miR-374a	low				high	1
miR-378 (422b)				high	not reported	0
miR-423-5p			low		not reported	0
miR-424	high				low	3
miR-497		high			low	1
miR-513	low				inconsistent	2
miR-563	low				high	1
miR-565	low				high	1
miR-575	low				high	1
miR-596			high		not reported	0
miR-598			high		not reported	0
miR-602	low				high	1
miR-623	low				high	1
miR-625			high		high	1
miR-625*			high		high	1
miR-628	low				high	1
miR-638	low				high	1
miR-744			low		not reported	0
miR-768-3p			high		not reported	0
miR-877			low		not reported	0
miR-92a				high	high	1
miR-93				high	high	3
miR-99b	**high**		**low**		low	1

Table 3.2. miRNA expression found in HPV-16/-18 cervical cancer samples by microarray. miRNAs which shown differences across studies are in grey cells

	Wang *et al.* (2008)	Wang *et al.* (2008) (*in vitro* HPV + raft tissues)	Pereira *et al.* (2010)	Li *et al.* (2011)	Lajer *et al.* (2012)
hsa-let-7a	high	high			
hsa-let-7b		**high**		**low**	
hsa-let-7c	high	high			
hsa-let-7d	high	high			
hsa-let-7e		high			

hsa-let-7f	high	high			
hsa-let-7g		high			
hsa-miR-100	low	low		low	low
hsa-miR-101		low			
hsa-miR-106a	high		high	high	high
hsa-miR-106b					high
hsa-miR-10a			high		
hsa-miR-10b		low		low	
hsa-miR-1224-5p					high
hsa-miR-125	low				
hsa-miR-125b		low		low	low
hsa-miR-126	low			low	
hsa-miR-127	low				
hsa-miR-130b					high
hsa-miR-132			high		
hsa-miR-133a	low				
hsa-miR-133b	low				
hsa-miR-142-5p			high		
hsa-miR-143	low	low	low		
hsa-miR-145	low	low	low	low	
hsa-miR-145*			low		
hsa-miR-146a	**high**	**low**			
hsa-miR-146b-5p		**low**			**high**
hsa-miR-148a	high		high		
hsa-miR-149					low
hsa-miR-155	high			high	high
hsa-miR-15a	high	high			
hsa-miR-15b	high			high	
hsa-miR-16	high		high	high	
hsa-miR-17-5p	high			high	
hsa-miR-17*					high
hsa-miR-181a		low			
hsa-miR-181b		low			
hsa-miR-181c	high				
hsa-miR-181d		low			
hsa-miR-182	high			high	
hsa-miR-183	high				
hsa-miR-185	high			high	high
hsa-miR-191*	**low**	**high**			
hsa-miR-195	**low**	**high**		**low**	**low**
hsa-miR-196a			high		
hsa-miR-197			high		
hsa-miR-198		high			
hsa-miR-199a	low	low	low		
hsa-miR-199a-3p				low	
hsa-miR-200b		high			
hsa-miR-200c		high			
hsa-miR-203		**high**	**low**		
hsa-miR-205	high	high	high		
hsa-miR-20a	high			high	
hsa-miR-20b	high			high	
hsa-miR-21	high	high			high
hsa-miR-214		low			
hsa-miR-218	low	low		low	
hsa-miR-22		low			
hsa-miR-221		**low**		**high**	
hsa-miR-222		**low**		**high**	
hsa-miR-223	high	high			

hsa-miR-224	high			high	
hsa-miR-23a	high	high			
hsa-miR-23b	high	high			
hsa-miR-24		low			
hsa-miR-25				high	
hsa-miR-26a	low	low	low		
hsa-miR-26b		high			
hsa-miR-27a		high	high		
hsa-miR-27b		high			
hsa-miR-29a	low	low	low	low	
hsa-miR-29c				low	
hsa-miR-301		low			
hsa-miR-302b			high		
hsa-miR-30a-5p		low			
hsa-miR-30b	low				
hsa-miR-30d		low			
hsa-miR-30e-5p		low			
hsa-miR-31				high	
hsa-miR-324-5p	low				
hsa-miR-339-5p					high
hsa-miR-34a		low			
hsa-miR-34c		low			
hsa-miR-374a		high			
hsa-miR-375				low	low
hsa-miR-422a	low				
hsa-miR-422b (378)	low				
hsa-miR-424	low	low		low	
hsa-miR-450	low				
hsa-miR-455	low				
hsa-miR-497					low
hsa-miR-512-5p			high		
hsa-miR-513		**high**	**low**		
hsa-miR-522*			high		
hsa-miR-563		high			
hsa-miR-565		high			
hsa-miR-574	low				
hsa-miR-575		high			
hsa-miR-602		high			
hsa-miR-617					low
hsa-miR-623		high			
hsa-miR-625					high
hsa-miR-628		high			
hsa-miR-638		high			
hsa-miR-92a				high	
hsa-miR-92b				high	
hsa-miR-93	high			high	high
hsa-miR-941					high
hsa-miR-99a	low		low	low	low
hsa-miR-99b		low			

Bovine papillomaviruses (BPVs) type 1 and 2 are recognised as causative agents of equine sarcoids (ES) (Nasir & Reid, 1999; Brandt *et al.*, 2011b; Wobeser *et al.*, 2012). There is scarce information of a third BPV (BPV-13) which may have a role in ES (Lunardi *et al.*, 2013b). The carcinogenic effects of BPV type -1 and -2

early proteins E5, E6 and E7 are the best characterised (Petti *et al.*, 1991; Borzacchiello *et al.*, 2006; Borzachiello *et al.*, 2008; Marchetti *et al.*, 2009; Yuan *et al.*, 2010a; Yuan *et al.*, 2010b; Yuan *et al.*, 2011c; Mosseri *et al.*, 2014). In fact, BPV-1 E5 alone is capable of transforming fibroblasts of mouse, human and equine (Bergman *et al.*, 1988; Petti & Ray, 2000; Yuan *et al.*, 2011b). BPV-1 E6 has the same transforming effect on equine fibroblasts as BPV-1 E5 and both contribute to the invasive phenotype of tumoural cells (Yuan *et al.*, 2011b). BPV-1 E7 on the other hand has a synergistic function and it has been demonstrated to be involved in anchorage-independent growth of transformed equine fibroblasts, together with BPV-1 E6 (Yuan *et al.*, 2011b). The roles of BPV-1 oncoproteins have been described in detail in Chapter 1, section 1.1.4.1.

Several studies have been conducted to identify the mechanisms by which BPV-1 and -2 oncoproteins induce and perpetuate tumoural transformation. Activation of the platelet-derived growth factor β receptor (PDGFβ-R) signaling pathway and the intracellular retention of major histocompatibility complex (MHC) class I, both driven by the BPV-1/-2 E5 oncoprotein, are the most studied cellular events (Petti *et al.*, 1991; Ashrafi *et al.*, 2002; Borzacchiello *et al.*, 2006; Borzachiello *et al.*, 2008; Marchetti *et al.*, 2009; Altamura *et al.*, 2013). Other recognised effects of BPV-1/-2 oncoproteins in ES are activation of p38 mitogen-activated protein kinases (p38-MAPK) (Yuan *et al.*, 2011c), intracellular mislocation of tumour suppressor p53 (Yuan *et al.*, 2008a), activation of p600, a member of the retinoblastoma family (Corteggio *et al.*, 2011), downregulation of toll-like receptor 4 (TLR4) (Yuan *et al.*, 2010a) and increased expression of forkhead box protein 3 (FOXP3) (Mahlmann *et al.*, 2014). A summary of the molecules/pathways altered by BPV-1/2 during oncogenesis is displayed in Table 3.3. However, to date, the mechanisms by which the virus activates or inhibits these pathways are still poorly understood.

A previous study investigated *in vitro* changes to the equine fibroblast transcriptome induced by BPV-1 using microarray expression profiling (Yuan *et al.*, 2008b). This demonstrated altered expression of genes involved in a range of processes including immunity and inflammation, cell adhesion, motility as well as cell cycle/cell proliferation and apoptosis. Currently, there is no information of miRNA expression changes in equine sarcoids.

Table 3.3. Molecules/pathways altered by BPV-1/-2 during tumoural transformation in equine sarcoids

Molecule / Pathway	Effect	BPV trigger	Source
Platelet derived growth factor beta receptor (PDGFβ-R)	Activation	BPV-1 E5	Petti *et al.* (1991); Borzachiello *et al.* (2008); Altamura *et al.* (2013)
Ras-mitogen-activated protein kinase pathway (Ras-MAPK)	Activation	BPV-1 E5?	Altamura *et al.* (2013)
PDGF-βR-PI3K-AKT signaling cascade	Activation	BPV-1 E5	Borzacchiello *et al.* (2009)
p38 mitogen-activated protein kinases (p38-MAPK)	Activation	BPV-1 E5?	Yuan *et al.* (2011c)
p53	Mislocation	BPV-1 E5	Yuan *et al.* (2008a)
p600 (part of the retinoblastoma family)	Activation	BPV-1 E7	DeMasi *et al.* (2005); Corteggio *et al.* (2011)
Matrix metalloproteinase MMP-1	Upregulation	BPV-1 E7/E5	Yuan *et al.* (2010b)
TLR4	Downregulation	BPV-1 E2 and E7	Yuan *et al.* (2008b); Yuan *et al.* (2010a)
MHC class I	Downregulation	BPV-1 E7	Marchetti *et al.* (2009)
Transcription factor forkhead box protein 3 (FOXP3)	Upregulation	? Possibly E5	Mahlmann *et al.* (2014)

Microarray means small size grid ('*mikro*', which means small in Greek), of probes orderly arranged ('*ad- redare'*, from Latin) (Trachtenberg *et al.*, 2012). The probes are affixed to a surface and they are made by small size single stranded DNA oligonucleotides. Utilising the affinity of single stranded cDNA to become double stranded, the microarray probes of DNA oligonucleotides bind to their complementary RNA sequences giving evidence for the presence of certain genes in the sample exposed to the probes. This hybridization is used to assess the presence of thousands of transcripts in an RNA sample. The basic components in a microarray analysis are the RNA sample being analysed (and amplification kit if initial amount of sample is too small), microarray chip, labelling kit, hybridization station and chambers, scanner for image capture and software for data analysis (Wang & Yang, 2010; Trachtenberg *et al.*, 2012). Microarray technology is widely used to examine gene expression and has been utilised in cervical cancer studies as a starting point to better understand the mechanisms of tumorigenesis induced by HPV (Wang *et al.*, 2014b).

The aim of this chapter was to use a novel approach to improve understanding of the molecular pathogenesis associated with BPV-1-induced tumorigenesis in sarcoids. The results presented here, obtained using microarray technology, identify cellular miRNAs that are differentially expressed in BPV-1-transformed

equine fibroblasts that may lead to BPV-1-induced tumorigenesis.

3.2. Materials and methods

3.2.1. Cell culture and RNA isolation

Equine primary fibroblasts (EqPalFs) and BPV-1 transformed cells (S6-2) have been previously described (Yuan *et al.*, 2008a). RNA was isolated using a phenol based method, a variation of the classic phenol/chloroform extraction and ethanol precipitation, with the use of Trizol® reagent. For details on the cell lines, cell culture conditions, RNA isolation, quantification and quality assessment of RNA, the reader is referred to Chapter 2, sections 2.2. (Cell culture) and 2.3. (Molecular biology). One RNA extraction was performed per cell line, and the RNA sample from each cell line was divided in 3 aliquots (technical replicates) for the microarray.

3.2.2. miRNA microarray

miRNA expression profiling was performed using the commercial array platform miRCURY LNA™ microRNA Array 7th Gen (Exiqon, Vedbaek Denmark) probed with total RNA isolated from EqPalF cells (normal equine fibroblast) and S6-2 (BPV-1 transformed fibroblast) cells. This array comprised 2,164 probes that contain human (1651 probes, 76.3%), viral miRNA sequences (146, 6.75%), predicted equine microRNA sequences derived from miRBase 19.0. (350, 16.2%) (Griffiths-Jones *et al.*, 2006) and SNORDs (17, 0.8%). A detailed description of the technique is given in Chapter 2 (Materials and methods) in section 2.4. (miRNA microarray) and a list of probes is included in the file 'master file microarray.xlsx' supplied in the DVD.

3.2.3. Validation of miRNA expression using real-time quantitative reverse transcription polymerase chain reaction (real-time qRT-PCR)

Eleven miRNAs were selected for validation of the microarray results. The list of miRNAs selected and qRT-PCR primer sequences are shown in the results section

(Table 3.7). miRNAs were quantified using SYBR® green qRT-PCR technology (Takara Bio Europe, St Germaine en Laye, France), however due to problems with non-specific amplification and low efficiency for some primer sets, a panel of miRNAs were examined using TaqMan® MicroRNA Assays (Life Technologies Ltd, Paisley, United Kingdom). Primers used for amplification were supplied as part of the kits with the exception of the 5' primers for SYBR green and these are listed in Table 3.4. Each amplification reaction was performed in triplicate and the average C_t value was used for subsequent analysis. Expression of each miRNA was normalised using RNU6 and differential expression was calculated using the $\Delta\Delta C_t$ method (Winer *et al.*, 1999b; da Silveira *et al.*, 2012; Buechli *et al.*, 2013). Further details of the qRT-PCR technique are included in Chapter 2, section 2.6. (Real-time quantitative reverse transcription polymerase chain reaction).

3.3. Results

3.3.1. Quality control of RNA samples

Triplicate samples of RNA isolated (three technical replicates per cell line) from EqPalF and S6-2 cells were subjected to NanoDrop® ND-1000 analysis to measure RNA concentration from optical density readings. In addition, on arrival at the microarray service provider, the RNA integrity number (RIN) of samples was assessed using the Bioanalyzer 2100. This tests the quality of RNA. The results from these measurements (concentration, optical densities and RIN value) are displayed in Table 3.4.

Table 3.4. Quality control and quantification of RNA extracted for microarray analysis. Values of concentration and optical density were obtained using NanoDrop® ND-1000 Spectrophotometer and RIN values were obtained using automated electrophoresis and flow cytometry (Bioanalyzer 2100, Agilent Technologies). Each replicate was obtained by dividing one RNA sample in three aliquots (technical replicates)

	Concentration (ng/µl)	Optical density (O.D.) 260/280	Optical density (O.D.) 260/230	RIN value
EqPalF-1	1455	1.88	2.22	10
EqPalF-2	1455	1.89	2.20	10
EqPalF-3	1455	1.83	2.16	10
S6-2-1	901	1.87	2.10	10
S6-2-2	901	1.88	2.10	10
S6-2-3	901	1.85	2.12	10

In addition, agarose gel electrophoresis of the samples revealed minimal degradation of the RNA samples as shown by the clear 18S and 28S bands in

Figure 3.1.

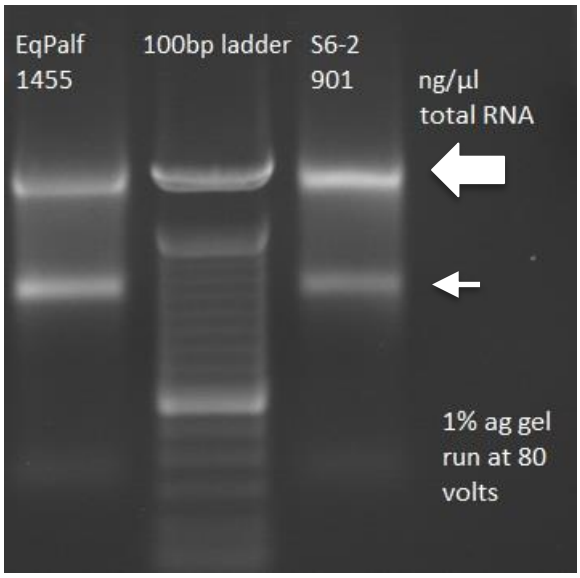


Figure 3.1. Agarose gel electrophoresis of RNA samples from EqPalF and S6-2 cells. Ribosomal bands 28S (bigger arrow) and 18S (small arrow) were well defined indicating limited degradation of the RNA

3.3.2. Microarray data

3.3.2.1. microRNAs expressed above background signal

Of the 2,164 probes present in the microarray, 492 probes shown signal intensity above the background. The background threshold was calculated for each individual microarray slide as 1.2 times the 25th percentile of the overall signal intensity of the slide (Exiqon, Vedbaek Denmark). Only miRNAs with intensities above threshold in more than one of the samples were included in the expression analysis dataset.

Principal component analysis (PCA) was used to investigate the relationship between the S6-2 and EqPalF microarray readings. PCA allow complex multi-dimensional datasets to be transformed to reveal underlying patterns of variation. Microarray expression data for the 492 probes was subjected to PCA

and clear separation was evident between the cell lines that containing BPV-1 (S6-2) and the control cell line (EqPalF) (see Figure 3.2.), with replicates of the same cell-type forming clusters. Replicates of S6-2 and EqPalF were discriminated by the first principal component, which accounted for more than 60% of the variation in the dataset. This indicated that the differences in microRNA expression profile of the two types of sample were pronounced and that a good level of biological replication was evident

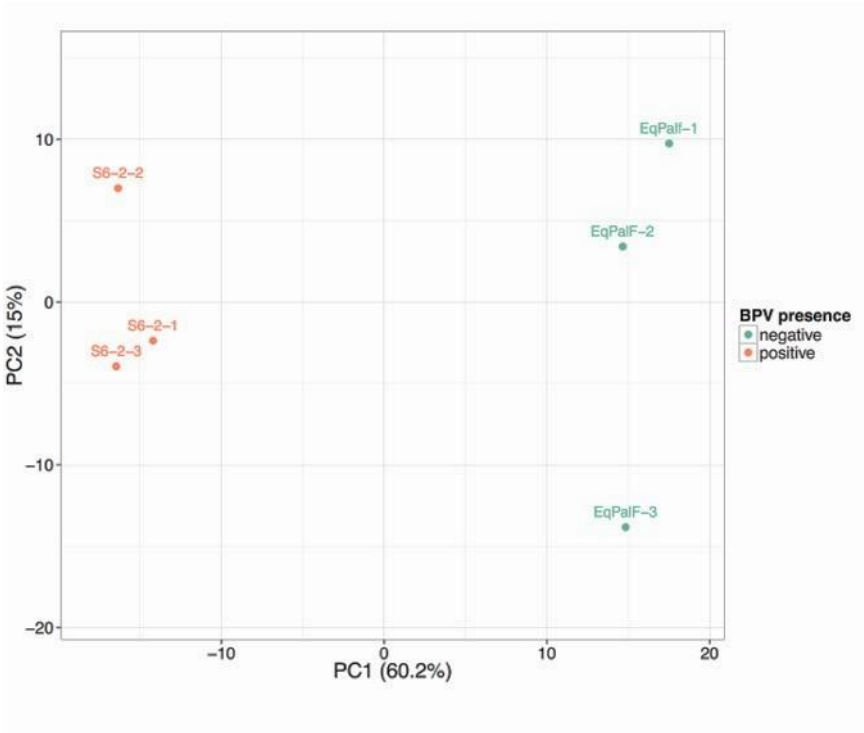


Figure 3.2. Principal component analysis (PCA) plot with 492 probes found above the background signal. Unit variance scaling is applied to rows; singular value decomposition with imputation is used to calculate principal components. The X and Y axes show the 1st and 2nd principal components, which explain 60.2 % and 15 % of the total variance, respectively. N = 6 data points (graph obtained using online tool Clustvis)

3.3.2.2. Data refinement

Of the 2,164 probes on the array, 492 were above the background signal (average signal intensity from three replicates) in both cell lines. Of those, 90 probes were excluded for the data analysis for various reasons, as detailed below. Twenty-three probes were duplicates, identified on the basis of their annotation or mature sequence name. The data from duplicated probes was averaged to produce a value for each individual mature miRNA sequence.

Another 67 probes were removed as they were included in the array design for quality control purposes. These included 19 human-targeted probes known to give false positive readings, 27 non-papillomavirus sequences, 9 human synthetic sequences and 12 small nucleolar RNA. Thus, in total, analysable data was generated representing 402 microRNA sequences. Of those, 124 were equine predicted sequences and 278 were human sequences. The contribution of each group of probes to the total is represented in Figure 3.3. and a summary of the data refinement is displayed in Figure 3.4. A detailed list of the miRNAs and sequences is included in the file 'master file microarray.xlsx' in the DVD supplied.

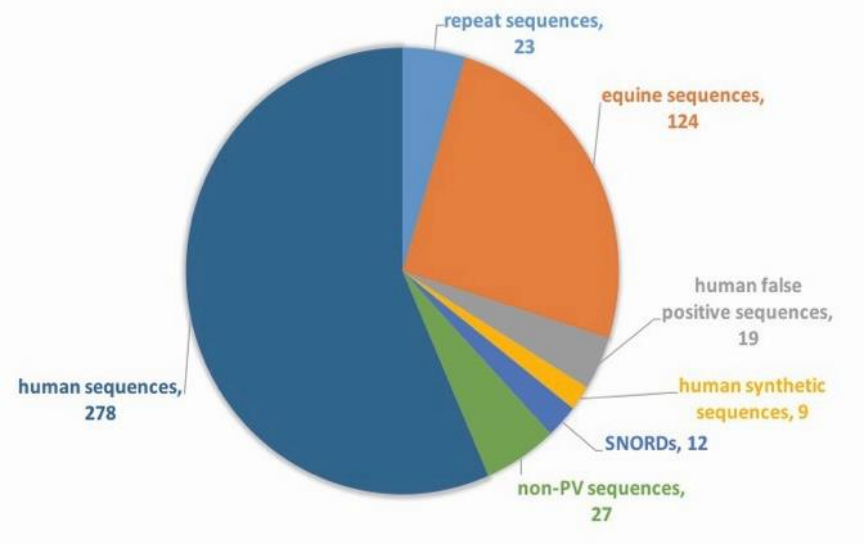


Figure 3.3. Pie chart displaying the distribution of the probes/sequences found in equine fibroblasts using miRNA microarray. 278 sequences were human mature miRNAs, 124 sequences were equine predicted sequences, 90 probes were excluded as they were repeats, false positives, synthetic sequences, non-PV viral sequences and SNORDs

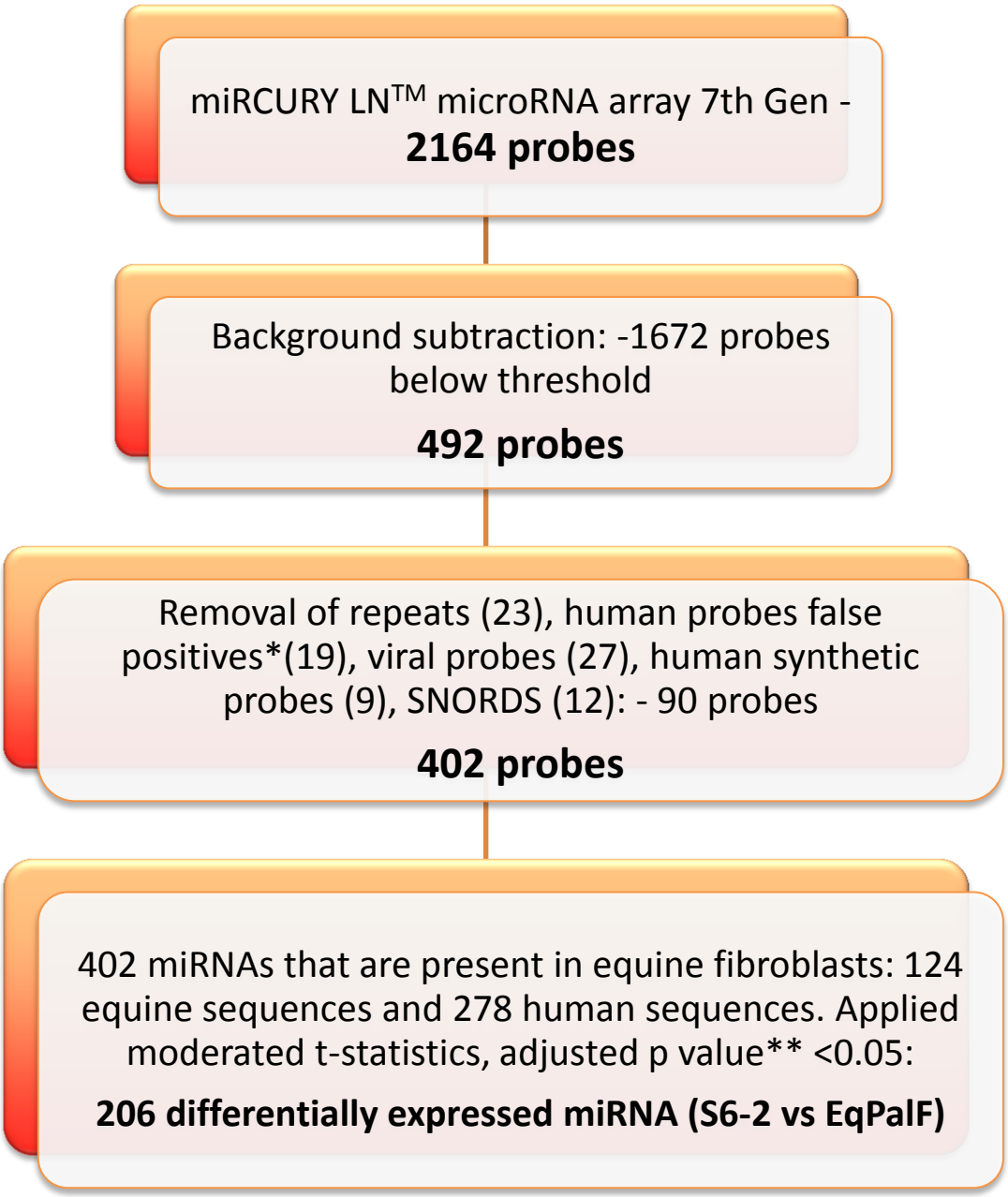


Figure 3.4. Summary of the steps of the analysis of differential expression of miRNAs between BPV-1 transformed cells (S6-2 cells) and EqPalFs using microarray. *human 'false-positive' probes are used as part of the quality control process and are discarded in the analysis. ** The Benjamini and Hochberg multiple testing adjustment method was applied on the data to obtain the adjusted p values

3.3.2.3. Differential expression of miRNAs

Just over 50% of the 402 sequences were found to be differentially expressed between the two cell types after applying moderated *t*-statistics (206, adjusted p value < 0.05 using the Benjamini and Hochberg multiple testing adjustment method). In total 62 (15%) miRNAs were found to be upregulated in S6-2 cells compared to EqPalFs, 144 (36%) miRNAs were downregulated and 196 (49%) were not significantly different ($p<0.05$). The distribution of up- and downregulated miRNAs is displayed in Figure 3.5. Log_2 fold change in the 206 differentially expressed miRNA ranged from -2.13 to 0.61 and signal intensity ranged from 5.81 to 13.53. The list of the top 30 downregulated miRNAs in S6-2 relative to EqPalF cells (ranked by FC) is displayed in Table 3.5. The list of top 30 upregulated miRNAs (ranked by FC) is presented in Table 3.6.

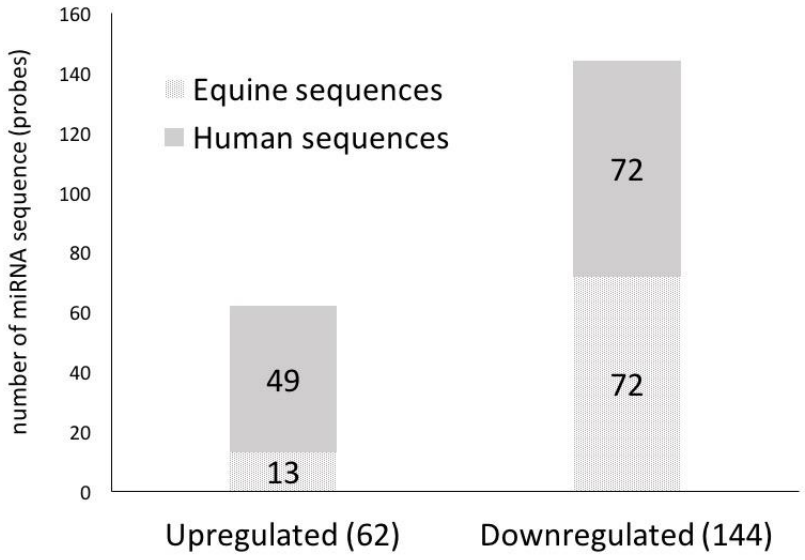


Figure 3.5. Distribution of up- and down-regulated miRNAs in S6-2 cells *vs*. EqPalFs. The majority of the dysregulated miRNAs were downregulated in S6-2 cells. In total 144 miRNA candidates were found downregulated (72 human sequences and 72 equine predicted sequences, 70% of the differentially expressed miRNAs) and 62 were found upregulated (49 human sequence and 13 equine predicted sequences, 30% of the differentially expressed miRNAs)

Table 3.5. Top 30 downregulated miRNA candidates in S6-2 cells compared to EqPalF cells (adjusted p value <0.05; ranked by FC)

Annotation	FC	AvgHy3	adj. p value
eca-miR-127	-4.38	5.92	1.34E-08
hsa-miR-4732-5p	-2.98	9.47	1.34E-08
eca-miR-132	-2.79	5.81	2.79E-07
eca-miR-181a	-2.41	8.21	3.65E-08
eca-miR-409-3p	-2.25	5.83	5.36E-07
hsa-miR-663a	-2.18	9.1	9.81E-08
hsa-miR-3158-5p	-2.03	6.51	2.79E-07
eca-miR-15b	-2.00	9.34	2.20E-07
eca-miR-106a/ eca-miR-17	-1.97	6.87	3.70E-07
hsa-miR-199a-3p	-1.95	9.2	5.02E-07
hsa-miR-675-5p	-1.91	6.63	3.16E-07
hsa-miR-455-3p	-1.88	7.27	3.13E-07
eca-miR-199b-5p	-1.86	8.69	2.20E-07
eca-miR-16	-1.86	9.63	1.40E-06
eca-miR-181b	-1.83	6.51	2.20E-07
eca-miR-199a-5p	-1.83	8.56	3.70E-07
eca-miR-23a	-1.80	11.98	7.60E-07
hsa-miR-106a-5p	-1.80	6.66	1.14E-06
eca-miR-24	-1.80	11.14	1.54E-05
eca-miR-93	-1.77	7.89	5.02E-07
hsa-let-7b-5p	-1.77	10.91	1.32E-06
eca-miR-23b	-1.76	12.37	5.53E-07
eca-miR-487b	-1.71	5.84	1.98E-06
eca-miR-214	-1.71	10.22	5.80E-07
eca-miR-27a	-1.68	8.12	1.14E-06
hsa-miR-181d	-1.68	6.93	1.74E-06
hsa-let-7i-5p	-1.66	8.42	1.31E-05
eca-miR-432	-1.66	6.64	5.77E-06
eca-miR-25	-1.66	7.27	4.54E-07
eca-miR-20a	-1.65	6.73	7.04E-07

Table 3.6. Top 30 upregulated miRNA candidates in S6-2 cells compared to EqPalF cells (adjusted p value <0.05) ranked by FC

Annotation	FC	AvgHy3	Adj.P.Val
eca-miR-1902	1.53	6.48	1.14E-05
hsa-miR-493-5p	1.48	8.99	8.20E-06
hsa-miR-4484	1.41	5.93	1.28E-05
hsa-miR-1285-3p	1.40	8.02	1.54E-05
hsa-miR-3121-5p	1.40	7.12	8.04E-06
hsa-miR-639	1.39	7.64	1.70E-04
eca-miR-193b	1.37	7.44	1.46E-05
hsa-miR-4742-3p	1.32	7.47	6.63E-05
eca-miR-1248	1.29	7.05	3.94E-05
eca-miR-99a	1.27	6.78	5.15E-05
hsa-miR-630	1.27	6.37	3.57E-04
hsa-miR-3676-5p	1.27	6.75	9.50E-05
hsa-miR-3161	1.25	6.18	1.87E-03
eca-miR-122	1.23	7.55	3.22E-02
hsa-miR-4301	1.23	11.81	1.39E-03
hsa-miR-1321	1.23	6.37	2.72E-04
eca-miR-187	1.22	6.82	7.58E-04
hsa-miR-514a-5p	1.21	6.32	3.61E-04
hsa-miR-5684	1.19	11.8	2.59E-03
hsa-miR-4791	1.19	6.9	1.51E-03
hsa-miR-377-5p	1.19	6.31	5.25E-03
hsa-miR-877-5p	1.19	7.15	4.86E-04
hsa-miR-636	1.19	6.5	2.25E-03
hsa-miR-4646-3p	1.18	6.31	3.90E-03
hsa-miR-4712-3p	1.17	6.61	4.15E-03
hsa-miR-4780	1.17	9.85	4.84E-03
hsa-miR-4475	1.17	7.92	2.11E-03
hsa-miR-4299	1.16	7.13	7.24E-04
hsa-miR-4682	1.17	6.68	5.00E-03
eca-miR-532-3p	1.17	6.31	2.85E-03

A two-way hierarchical clustering of miRNAs and samples showed that the EqPalF replicates are closer to each other than they are to S6-2 and equally the replicates of S6-2 are closer to each other than they are to EqPalF (Figure 3.6). Additionally, the clustering indicates, based on the distance of the tree branches, that there is a clear discrimination between the two cell lines. The heat map also highlights that the majority of differentially expressed miRNAs in S6-2 cells are downregulated.

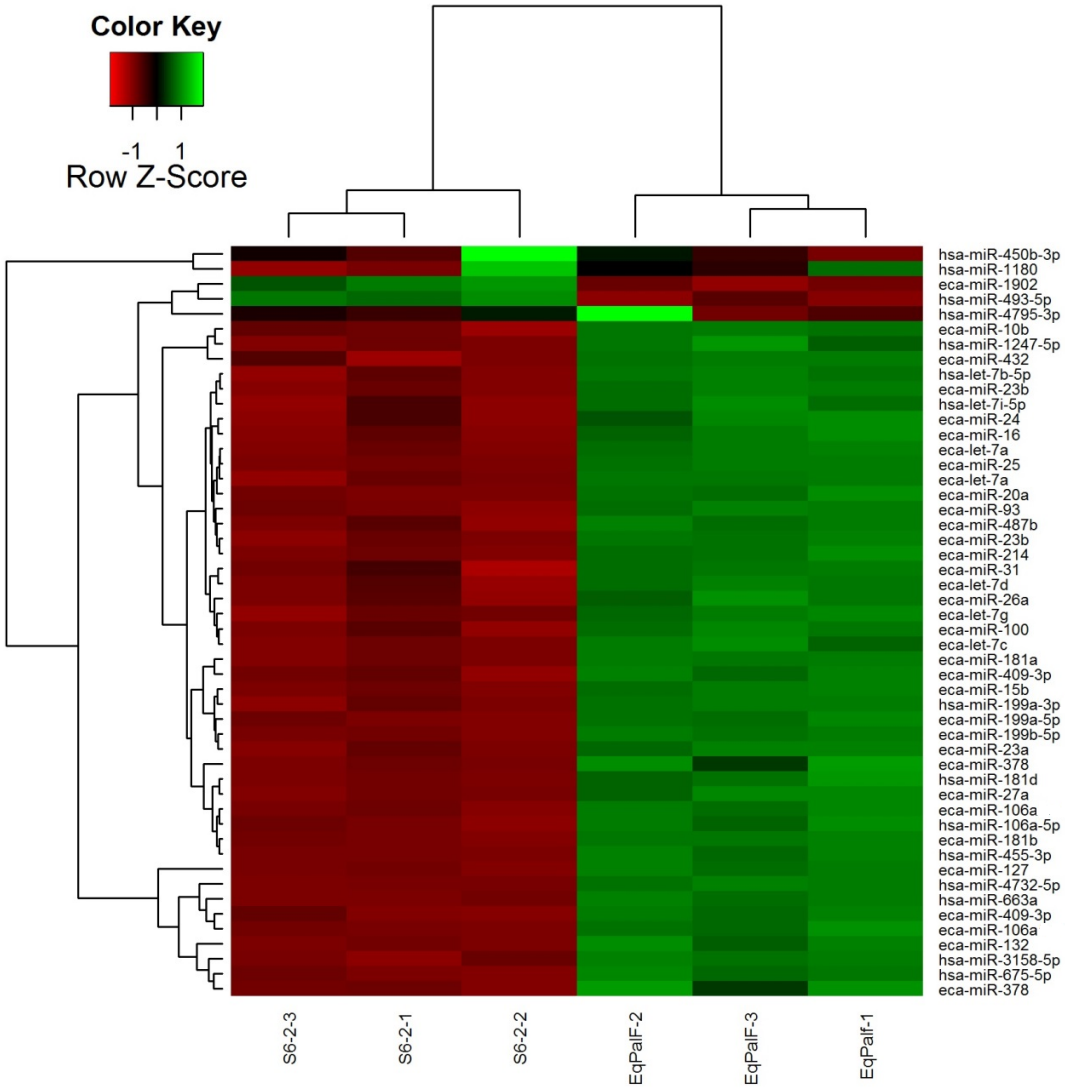


Figure 3.6. Heat map representing the top 50 miRNAs (with the highest standard deviation) out of 206 differentially expressed miRNAs between EqPalF and S6-2 cells using microarray. The complete-linkage method was used to define clusters together with a euclidean distance measure. Each row represents a

microRNA and each column represents a sample. The microRNA clustering tree is shown on the left. The colour scale illustrates the relative expression level of microRNAs (each replicate compared to the reference). Red represents an expression level below the reference channel and green represents expression higher than the reference

3.3.3. Real-time qRT-PCR data

To validate the S6-2 versus EqPalF microarray results, 11 differentially expressed miRNAs were assessed by qRT-PCR using RNA prepared from three biological replicates of the same cell lines. RNA preparations were obtained from other batches of cells, different from those used in the microarray analysis, and were extracted using the same method as described above. The miRNAs were chosen based on two of the following thresholds advised by Exiqon: absolute fold-change of ≥2, microarray signal intensity between 7.5 and 14.5 and biological significance, i.e. previously shown to be implicated in human papillomavirus-associated cancer. The list of miRNAs selected for validation with the results from the microarray and sequence is displayed in Table 3.7.

Table 3.7. List of 11 miRNAs selected for real-time qRT-PCR validation of microarray data. Four criteria were used to select the miRNAs differentially expressed between BPV-1 transformed equine fibroblasts (S6-2 cells) and normal equine fibroblasts (EqPalF cells) with significance of $p<0.05$ (adjusted p value); fold change between the two cell lines (FC) outwith 2.0 and -2.0, signal intensity on the microarray (AvgHy3) between 7.5 and 14.5 and biological significance (b.s.) (previously found dysregulated in cancer and or HPV induced cancer)

Annotation	FC	AvgHy3	Adj. p value	Criteria	Sequence	miRBase accession number	Biological significance (b.s.)
eca-miR-15b	-2.0	9.34	2.20E-07	Hy3, FC criteria and b.s.	TAGCAGCACATCATGGTTTACA	MIMAT0012954	Reported to be upregulated in HR-HPV (Wang *et al.*, 2008; Li *et al.*, 2011; Lajer *et al.*, 2012) and downregulated in glioma cells (Sun *et al.*, 2014)
hsa-miR-663a	-2.17	9.1	9.81E-08	Hy3, FC criteria and b.s.	AGGCGGGGCGCCGCGGGACCGC	MIMAT0003326	Promotes switch to abnormal phenotype in proliferative cardiac disease model (Liu, 2013).
eca-miR-181a	-2.41	8.21	3.65E-08	Hy3, FC criteria and b.s.	AACATTCAACGCTGTCGGTGAGT	MIMAT0013178	High in HPV + monolayer culture but low in HPV + raft tissues using microarray (Wang *et al.*, 2008); low in HPV-16 cell line using microarray and qRT-PCR (Wald *et al.*, 2011), low in HPV-16/18 cell line using microarray and qRT-PCR (Lajer *et al.*, 2012). Encoded in ECA25, region associated with equine sarcoids (Jandova *et al.*, 2012).
hsa-miR-4732-5p	-2.97	9.47	1.34E-08	Hy3, FC criteria and b.s.	TGTAGAGCAGGGAGCAGGAAGCT	MIMAT0019855	Regulates abundance of tumour suppressor p53 (Pouladi *et al.*, 2013).
eca-miR-127 (hsa-miR-127-3p)	-4.38	5.92	1.34E-08	FC criteria and b.s	TCGGATCCGTCTGAGCTTGGCT	MIMAT0013126	Downregulated in HPV induced cancer (Zheng & Wang, 2011).
eca-miR-409-3p	-2.25	5.83	5.36E-07	FC criteria and b.s	GAATGTTGCTCGGTGAACCCCT	MIMAT0013151	Downregulated in HPV-positive cells in pharyngeal SCC (Lajer *et al.*, 2011). It suppresses cancer cell growth in breast cancer (Zhang *et al.*, 2016)
eca-miR-106a / eca-miR-17	1.97	6.87	3.70E-07	FC criteria and b.s	CAAAGTGCTTACAGTGCAGGTAG	MIMAT0013194	Upregulated in HPV16 cervical cancer (Zheng and Wang, 2011; Li et al., 2011)
eca-miR-132 (hsa-miR-132-3p)	-2.79	5.81	2.79E-07	FC criteria and b.s	TAACAGTCTACAGCCATGGTCG	MIMAT0013021	Upregulated in HPV cervical cancer (Pereira *et al.*, 2010)
hsa-let-7b-5p	-1.77	10.91	1.32E-06	Hy3 criteria and b.s.	TGAGGTAGTAGGTTGTGTGGTT	MIMAT0000063	Overexpression of let-7b inhibits invasion and migration in gastric cancer (Han *et al.*, 2015)

(Table 3.7 Continuation)

Annotation	FC	AvgHy3	Adj. p value	Criteria	Sequence	miRBase accession number	Biological significance (b.s.)
eca-miR-23b	-1.77	12.37	5.53E-07	Hy3 criteria and b.s.	ATCACATTGCCAGGGATTACC	MIMAT0013113	Reported to be downregulated in HPV-16 via E6 (Yeung *et al.*, 2011). Encoded in ECA23, region associated with equine sarcoids (Jandova et al., 2012).
eca-miR-143	-1.31	10.1	1.79E-04	Hy3 criteria and b.s.	TGAGATGAAGCACTGTAGCTC	MIMAT0013063	Lajer et al., 2012 - suppresses cell growth in HPV cancer. It is downregulated in hpv infected cells. Cluster 143/145 frequently mentioned as tumor suppressors. Mentioned in the majority of the HPV cancer related papers (summary Zheng and Wang, 2011)

Real-time quantification showed that 6 of the 11 microRNAs tested had significantly higher expression in EqPalF than in S6-2 cells, as expected (let-7b, miR-17, miR-23b, miR-132, miR-143, miR-181a) (Figure 3.7.), confirming the results from the microarray. Pearson's correlation coefficient between microarray and qPCR data was 0.68 (Figure 3.8.). One miRNA (miR-663a) showed no significant difference between the cell types, whilst two miRNAs had an inverse expression pattern to the microarray data (miR-15b and miR-127). Two miRNAs (miR-409 and miR-4732) could not be quantified due to low efficiency and/or non-specific amplification.

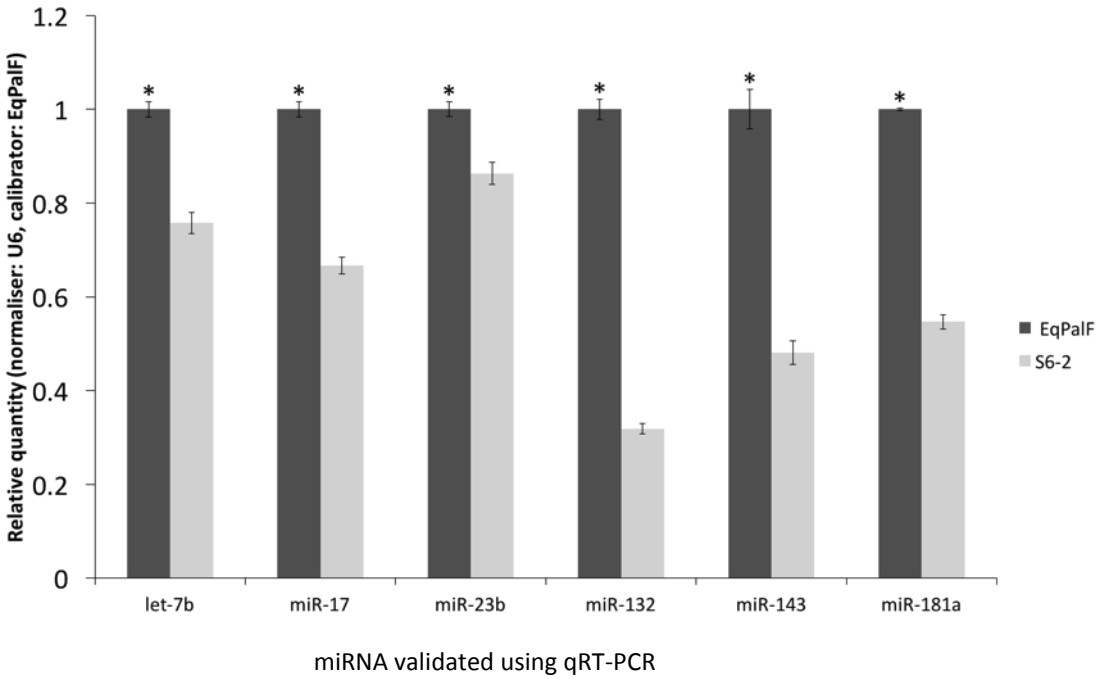


Figure 3.7. Relative quantification of the six miRNAs that validated microarray data using real-time qRT-PCR. The mean expression value of target miRNA was normalised against the expression of endogenous control U6 snRNA and values were calibrated against expression levels of calibrator (EqPalF). Error bars represent the standard error of the mean. Statistical significance is represented by asterisks, *P < 0.01

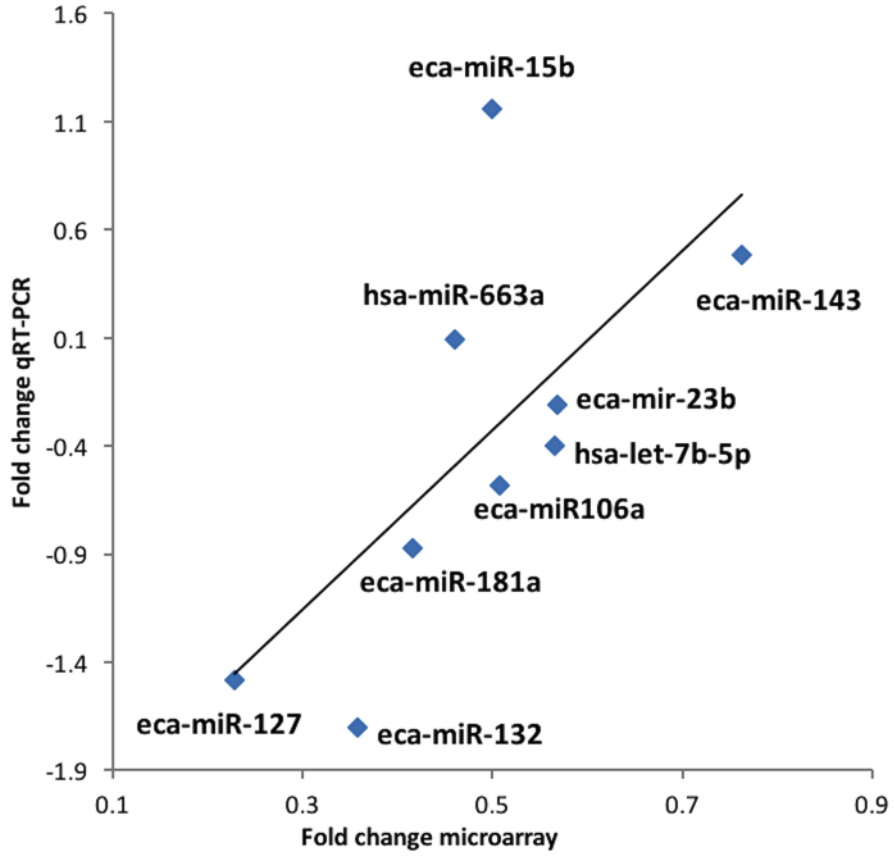


Figure 3.8. Correlation between microarray and qRT-PCR data. The differential expression of miRNAs in control equine fibroblasts (EqPalF) and BPV-1 transformed equine fibroblasts (S6-2 cells) was calculated by microarray (*x*-axis, S6-2/EqPalF ratio) and qRT-PCR (*y*-axis, log2 fold-change of S6-2 versus EqPalF)

3.4. Discussion

In total, 206 miRNAs were found to be significantly dysregulated in BPV-transformed equine primary fibroblasts when compared to control primary fibroblast using miRNA microarray (adjusted $p<0.05$). The results indicated a clear difference in miRNA expression between EqPalF and S-62 cells: replicates of the same cell line are located in the same region of the PCA plot and there is a clear separation between the two cell lines within the PCA plot. The strong correlation between microarray results and qRT-PCR is good supporting evidence indicating that the microarray results are a reliable estimate of relative miRNA expression.

Research and interest in the role of microRNAs in virus-host interactions has increased in recent years (Gupta *et al.*, 2012; Buggele & Horvath, 2013; Chang

et al., 2013). Viruses are able to remodel cellular miRNA expression networks impacting on the pathogenesis of virally induced cancers (Greco *et al.*, 2011; Yeung *et al.*, 2011; Wang *et al.*, 2014b). To date, there are few studies on equine miRNAs and no studies have investigated the impact of viral infection on miRNA expression in equine cells. For further details on equine miRNA studies the reader is advised to refer to Chapter 1, section 1.2.5.

miRNAs are highly conserved across species (Lee *et al.*, 2007) and phylogenetic conservation of the mature sequence was utilised in the first report on equine miRNAs (Zhou *et al.*, 2009). Using an integrated comparative genomic approach to the human genome and other animal genomes, a total of 354 mature sequences of miRNA candidates were computationally predicted in the horse (Zhou *et al.*, 2009). The results from the microarray comparing miRNA expression levels in EqPalF and S6-2 cells confirmed that 124 of these predicted miRNA are indeed encoded within the equine genome and expressed in equine fibroblasts.

Pathway analysis to group the candidate miRNAs into functional processes was attempted however, due to the low number of mRNA in the transcriptome dataset (Yuan *et al.*, 2008b) and the low number of microRNAs from the microarray identified in the Ingenuity Pathway Analysis database, the analysis was insufficient to provide information on canonical pathways (http://www.ingenuity.com/products/ipa).

An interesting observation in our study was that most of the miRNAs (70%) were downregulated in transformed cells. Pereira *et al.* (2010) assessed samples of human cervical cancer linked to HPV-16/-18 and found a general pattern of miRNA downregulation in tumour samples. However, it needs to be considered that the scope of the miRNA microarray performed in the current study was limited to 402 mature miRNA candidates so there is the possibility of a bias in the data due to the probe selection and the inherent problem of working with human sequences and equine predicted sequences. A general dowregulation of cellular miRNAs has also been observed in other cancers and some of the miRNAs found downregulated in S6-2 cells by both microarray and qRT-PCR in this study have been previously found to be downregulated in HPV-induced cancers either in cell culture or from *in vivo* tumours. These include let-7b, miR-23b and miR-

143 (Wang *et al.*, 2008; Pereira *et al.*, 2010; Li *et al.*, 2011; Yeung *et al.*, 2011; Liu *et al.*, 2012) and the data shown here indicate that these same miRNAs can be downregulated by PV in different host species.

Real-time qRT-PCR successfully validated differential expression of six of 11 selected microRNAs (let-7b, miR-17/miR-106a, miR-23b, miR-132, miR-143, miR-181a). For two of the selected miRNAs PCR reactions could not be optimised due to low efficiency and/or non-specific amplification (miR-4732-5p and miR-409-3p). miR-663a was shown to be significantly downregulated by microarray analysis (fold change of -2.17), however real-time qRT-PCR showed no difference in expression. Similarly, it was not possible to reproduce the results for miR-15b and miR-127 as they showed an inverse relationship by quantitative PCR compared to the microarray. No non-specific amplification was recorded for those miRNAs and the differences between both methods could have been related with presence of isomiR forms not detected by the microarray.

In this study let-7b was significantly downregulated by BPV-1 expression. In most human cancers, let-7b is considered to function as a tumour suppressor inhibiting cancer cell proliferation and tumour growth (Li *et al.*, 2011; Hung *et al.*, 2015; Yu *et al.*, 2015). miRNAs from the let-7 family reduce cancer cell proliferation and tumour development by controlling the expression of a number of oncogenes, such as *RAS* and *MYC* (Gomez-Gomez *et al.*, 2013). Let-7b has been shown to be downregulated in HPV-16/-18 infected keratinocytes, HPV-16 squamous cell carcinomas and HPV-18 cell cultures (Wang *et al.*, 2008; Li *et al.*, 2011). Since let-7b can function as a tumour suppressor, the downregulation of let-7b expression by BPV-1 may contribute to the tumoural transformation of equine cells.

miR-17/106a expression was also significantly downregulated in S6-2 cells. miR-17 has been reported to promote cell proliferation, to reduce apoptosis and to participate in the immune response to HPV-induced cervical cancer (Gomez-Gomez *et al.*, 2013). Wang *et al.* (2008) and Li *et al.* (2011) have shown that miR-17 is upregulated in HPV 16 cervical tissues when compared to normal cervix. The observed downregulation in S6-2 cells suggests that miR-17 can be modulated differently by distinct papillomaviruses.

Analysis of miR-23b by real-time qRT-PCR demonstrated that this miRNA is also downregulated. miR-23b has been shown to be involved in a wide range of cellular processes (Zhu *et al.*, 2012; Iaconetti *et al.*, 2015) and aberrant expression of miR-23b has been reported in many types of cancers (Goto *et al.*, 2014; Ma *et al.*, 2014; Avci *et al.*, 2015). Zhang *et al.* (2011a) showed that miR-23b suppressed invasion and metastasis both *in vitro* and *in vivo* and an independent study shown that this miRNA is linked to metastasis and invasiveness (Yeung *et al.*, 2011). HPV-16/-18 E6 protein has been reported to downregulate expression of miR-23b via expression of urokinase type plasminogen activator (uPA) and through a p53 dependent mechanism (Yeung *et al.*, 2011; Honegger *et al.*, 2015) although in a different study it has been reported to be upregulated in HPV-positive cell types (Wang et al. 2008). We have previously shown that BPV-1 oncogene E5 and E6 induced invasion of equine cells (Yuan *et al.*, 2008b) and the downregulation of miR-23b by viral oncogenes may represent a pathway by which this occurs. miR-23b has been predicted to be encoded within ECA 23 in a chromosomal region linked to the development of sarcoids (Jandova *et al.*, 2012).

The results from microarray and qRT-PCR show that miR-132 is significantly downregulated in S6-2 cells compared with EqPalF cells. This miRNA has been shown to be upregulated in HPV cervical cancer (Pereira *et al.*, 2010) and gastric cancer (Liu *et al.*, 2014). However, in other types of cancer such as breast cancer, osteosarcoma and pancreatic cancer miR-132 has been found to be downregulated (Zhang *et al.*, 2011b; Li *et al.*, 2013; Yang *et al.*, 2013). It is likely that the mechanisms of dysregulation of miR-132 by BPV-1 in ES may differ to those present in HPV induced cancer.

miR-143 is downregulated by BPV-1. miR-143 has been shown to be dysregulated by HPV-16 leading to inability to stop proliferation in cervical cancer via the p53 pathway. It has been claimed that miR-143 is an anti-tumoural miRNA and is consistently shown to be downregulated in HPV-induced cancer (Wang *et al.*, 2008; Witten *et al.*, 2010; Li *et al.*, 2011; Lajer *et al.*, 2012; Liu *et al.*, 2012). In the human genome miR-143 is encoded in a region that is deleted in prostate cancer (Pereira *et al.*, 2010).

Wald *et al.* (2011) showed that the presence of the oncogene HPV-16 E6 and E7 results in down-regulation of miR-181a in human cells. Roles attributed to this miRNA are inhibition of virus infection and replication in porcine reproductive and respiratory syndrome virus, activation of anti-inflammatory pathways, a tumour-suppressor role and as a prognostic biomarker in leukaemia patients and non-small cell lung cancer patients (Gao *et al.*, 2010; Schwind *et al.*, 2010; Fei *et al.*, 2012; Guo *et al.*, 2013; Xie *et al.*, 2013). This miRNA has been predicted to be encoded in a chromosomal region that has been linked to the development of sarcoids in Warmblood horses (Jandova *et al.*, 2012).

miR-663a was found to be significantly downregulated in S6-2 cells using microarray but this could not be reproduced with the use of qRT-PCR, as both EqPalF and S6-2 cells showed the same level of expression with the latter. SYBR green technology, rather than Taqman probe, was utilised for quantification of this miRNA so there may be the possibility of amplification of a highly similar miRNA in EqPalF cells.

miR-15b was shown to be downregulated in S6-2 cells in the data from the microarray although real-time qRT-PCR showed that in this cell type there was upregulation of this miRNA. This may have been related to the presence of isomiRs, which may have been quantified during qRT-PCR but not detected by the microarray probes. In HR-HPV induced cancer, miR-15b has been found to be consistently upregulated in cervical cancer tissue, HPV-16 squamous cell carcinoma lesions and in HPV-16 cell culture plates (Wang *et al.*, 2008; Li *et al.*, 2011; Lajer *et al.*, 2012; Wang *et al.*, 2014b). Further tests would be required to ascertain the level of miR-15b in BPV transformed cells and the role of this miRNA in ES.

miR-127 was found to be downregulated in the microarray data and upregulated in the qRT-PCR data in S6-2 cells. Quantification of this miRNA was performed using SYBR green technology using a predicted miRNA sequence for this miRNA. It is therefore possible that the amount of miRNA may have been overestimated as the specificity of the SYBR green method is lower compared to other qRT-PCR technologies and the possibility of isomiRs being picked up during qRT-PCR and not by the microarray probes. miR-127 has been found to be downregulated in cervical cancer tissues (Wang *et al.*, 2008). Similar to the finding with miR-15b,

further tests would need to be performed to ascertain the level and potential function of this miRNA in BPV transformed cells.

In summary, this is the first report of equine miRNAs that are differentially regulated by BPV-1. Future work using high throughput sequencing will be necessary to verify the equine miRNA sequences identified here, however this study should provide a basis for further studies investigating detailed roles on miRNAs in BPV-1 induced transformation and in the pathogenesis of equine sarcoid tumours.

Chapter 4. Identification of microRNAs in equine primary fibroblasts using high-throughput sequencing

4.1. Introduction

MicroRNAs (miRNAs) are single-stranded RNA molecules 18-23 nucleotides in length that regulate gene expression at the post-transcriptional level (Ambros, 2004). Understanding of microRNA biogenesis has challenged the central dogma of biology, since these molecules are the proof that many genes have a regulatory rather than protein coding function (Kim *et al.*, 2009). Through partial sequence complementarity, mature microRNAs bind to their target mRNA, leading to direct mRNA degradation, acceleration of mRNA degradation via deadenylation and/or suppression of protein translation (Filipowicz *et al.*, 2008). The same microRNA can target multiple mRNAs and the same mRNA can be the target of more than one mature microRNA. This mechanism enables microRNAs to 'switch on' and 'switch off' pathways involved in the cell cycle, proliferation, apoptosis, and senescence (Lu *et al.*, 2005; Rosenfeld *et al.*, 2008; Wang *et al.*, 2011; Wu *et al.*, 2014). Dysfunctional expression of microRNAs has been linked to multiple pathological mechanisms including cancer (Calin & Croce, 2006; Venugopal *et al.*, 2010; Kim *et al.*, 2011a; Baek *et al.*, 2015). Through activating or repressing certain cellular processes, some microRNAs can have an oncogenic effect or tumour suppressor effect (Huang *et al.*, 2015; Tutar *et al.*, 2015).

The horse genome (*Equus caballus*) was first assembled in 2007 (EquCab2.0, GenBank assembly accession reference GCA_000002305.1, http://www.ncbi.nlm.nih.gov/assembly/286598/) and since then a small number of studies have started to explore microRNA profiling in the horse. Currently there are 35,828 miRNAs in the official database miRBase 20.0 (www.mirbase.org) (Kozomara & Griffiths-Jones, 2014) of which 711 are annotated as equine miRNA precursors, which encode 690 mature miRNAs. More than half of the equine miRNAs in the database – 359 sequences - are *in silico* predicted (Zhou *et al.*, 2009), and prior to this study there was no data available to confirm that these are expressed miRNAs in the horse. The rest – 331 sequences – were obtained using high throughput sequencing (HTS) from a single testicular sample of a horse as part of a genetic study undertaken in large bats (*Eptesicus fuscus*), whereby the miRNA profiles from testicular samples from one bat, one dog and one horse were compared (Platt *et al.*, 2014). Although these

331 sequences are included in the official database, no details regarding data validation, such as real-time, reverse transcription quantitative PCR (real-time qRT-qPCR), are reported in the paper. Nonetheless, there have been studies attempting to describe miRNAs in the horse using HTS (Das *et al.*, 2013; Desjardin *et al.*, 2014; Kim *et al.*, 2014). These are summarised in Table 4.1.

Table 4.1. Summary of experimental design and main finding of previous studies on equine miRNAs identified by HTS

	Das *et al.* (2013)	Desjardin *et al.* (2014)	Kim *et al.* (2014)	Platt *et al.* (2014)
Tissue / sample	Semen	Bone and cartilage	Liver, colon and muscle	Testicular sample
Number of samples	8	3 for each condition and tissue	8 from each tissue	1
Type of samples	Healthy	Healthy *vs.* with osteochondritis	Healthy	Not reported
Validation	Not reported	qRT-PCR Taqman: 6 miRNA tested. 2 failed	Not reported	Not reported
Number of mature miRNAs with annotation	91	371	292	1
Suggested novel miRNAs	0	299	329	330
Remarks	6 miRNAs most abundant: miR-34b, miR-34c, miR-191, miR-223, miR-1248, miR-1905c	49 miRNAs in cartilage and 41 in bone differentially expressed between conditions (5 and 8 known respectively)	Specific mature miRNAs: - colon: 99 - muscle: 36 - liver: 3	Phylogenetic-comparative study. No detailed information on equine miRNAs given
Included in miRBase	No	No	No	Yes

Although empirical evidence of expression of some equine microRNAs has been documented, this has not been validated further and there is inconsistency in length and nucleotide composition of miRNA candidates across theses studies. An example of this disparity is eca-let-7a. This mature miRNA, with the annotation identifier MIMAT0012979 in the official database, has been predicted to be encoded in the horse in two genomic locations, chromosome 7 (from a precursor eca-let-7a) and chromosome 28 (from the precursor eca-let-7a-1). This mature miRNA is listed in Desjardin et al. (2014) and Kim et al. (2014) but both the name and nucleotide composition are dissimilar to those listed in miRBase. In Figure 4.1 the alignment of the three sequences shows that there is one nucleotide missing in the 5' end of the miRNA reported by Desjardin et al.

(2014), and 1 and 4 extra nucleotides at the 5' and 3' ends respectively in the miRNA reported by Kim et al. (2014) when compared to the miRNA annotated in miRBase.

```
Desjardin et al.       - eca-let-7a-5p     -GAGGUAGUAGGUUGUAUAGUU----
Kim et al.             - eca-let-7a        UGAGGUAGUAGGUUGUAUAGUUUAGA
miRBase (MIMAT0012979)- eca-let-7a        UGAGGUAGUAGGUUGUAUAGUU----
                                           ********************
```

Figure 4.1. Alignment of the mature miRNA eca-let-7a (MIMAT0012979) with the sequences of this miRNA reported by two HTS studies. It is noticeable that the length of the miRNA is different in the three datasets and additions and deletions are evident (highlighted in pink). This affects the definition of the seed region (from the 2nd to the 7th nucleotide from the 5' end) and consequently any target search performed using the seed region (Manzano *et al.*, 2015)

Variations in miRNA length or sequence can give rise to isomiRs (Morin *et al.*, 2008; Guo & Chen, 2014), which can make precise naming and comparison across studies difficult. IsomiRs can arise due to slight variations in miRNA precursor trimming at the 5' or the 3' end by Drosha and/or Dicer and also by addition of nucleotides to the 3' end of miRNAs (Neilsen *et al.*, 2012). The former most likely explains the extra sequence in eca-let-7a described by Kim *et al.* (2014) shown in Figure 4.1 as the four nucleotides 'UAGA' are part of the stem loop of the predicted precursor eca-let-7a (MI0012731) deposited in miRBase.

There is therefore a need for improved characterisation of miRNAs in the horse. In addition, no studies to date have sequenced miRNAs present in equine skin or fibroblast cells, both of which are relevant to the study of equine sarcoids. Thus, a goal of the present study is to produce a detailed catalogue of mature sequences expressed in equine fibroblast cells.

The comparative study of microRNAs using microarray technology (Chapter 3) supports the view that there is a high degree of sequence conservation across species. The probes utilised in the microarray were designed using mainly human microRNA sequences (1651 probes, 76.3%) and equine *in silico* predicted miRNA sequences from Zhou et al. (2009) (350 probes, 16.3%), and yet 22.7% of the probes (492) successfully hybridised (see Chapter 3, section 3.3.2.2. Data refinement). Microarray analysis is a relatively inexpensive platform that

requires very small amounts of RNA and allows a large number of genes to be interrogated from the same sample (Wang & Yang, 2010; Trachtenberg *et al.*, 2012). However, the recent introduction of HTS is beginning to supercede microarray technology. HTS allows absolute quantification of microRNA levels and this is reflected in its much higher dynamic range than microarray (Metzker, 2010). Another advantage of HTS over microarray analysis is that it does not depend on previous knowledge of the sequences of the miRNAs and is therefore an excellent tool for discovery of novel miRNAs (Wang *et al.*, 2009; Witten *et al.*, 2010). Sequence-based technologies have become more popular to profile microRNAs in different tissues. The drawback of HTS is mainly technical; compared to microarray analysis, more RNA is needed and this has to be of high quality for small RNA library preparation and sequencing. In addition, the output data requires greater analysis involving mapping to the genome of interest and removal of non-miRNA sequences to focus on miRNA expression.

Fibroblasts are found in a wide range of animals and primary fibroblasts represent a good example of a control for *in vitro* models of diseases such as Parkinson's disease and cardiomyopathies (Auburger *et al.*, 2012; Castaldo *et al.*, 2013). EqPalFs are equine primary fibroblasts and, in previous work at the University of Glasgow (Yuan *et al.*, 2008a), were used for BPV-1 transformation to create a robust *in vitro* model of equine sarcoids, As the focus of my PhD project was on miRNA expression in equine sarcoids, a first step in this study was to establish baseline expression of miRNAs in horse fibroblast cells. The main aim of the work detailed in this chapter was to use high-throughput sequencing technology (HTS) to obtain a detailed profile of microRNAs expressed in equine primary fibroblasts (EqPalFs).

4.2. Materials and methods

4.2.1. RNA extraction, quantitation and assessment of RNA integrity (RIN)

Chapter 2 (General materials and methods) contains details of cell culture, RNA extraction and RNA quality/quantity assessment. In summary, RNA was prepared using a phenol extraction method (with Trizol®) from equine primary fibroblasts

(EqPalF). Quantity and integrity were assessed and only when the sample had RIN values greater than 9.0 was it utilised for sequencing. The RNA obtained was divided into three samples for three library preparations.

4.2.2. Library construction, cluster generation and sequencing

A detailed explanation of library construction, cluster generation and sequencing is given in Chapter 2, Materials and Methods, section 2.5.1. In summary, libraries were prepared in triplicate using TruSeq® Small RNA (Life Technologies). Library preparation specific for miRNAs involves adaptor ligation and size selection so that only miRNAs are selected for further sequencing. After library construction, cluster generation on solid phase was performed followed by massive parallel sequencing in the Genomic Analyzer IIx from Illumina® (GA-IIx).

4.2.3. Data analysis

Data analysis of the HTS data is fully explained in Chapter 2 Materials and methods, section 2.5.2. A diagram explaining the steps in the pipeline can be found Chapter 2 (Figure 2.1). A summary is displayed in Figure 4.2.

In brief, the files obtained from the GA-IIx sequencer were pre-processed (removal of adaptor sequences), fed into miRDeep 2, software designed for microRNA identification and discovery. Sequences were mapped against the equine genome (EqCab.2.0) and the list of miRNA from the database (www.miRBase.org), following removal of redundant sequences and sequences smaller than 18 nucleotides in length, a list of equine predicted miRNAs was generated *de novo* based on RNA folding energy potential. Data was then harmonised using nucleotide composition and genomic location. Tolerance for nucleotide composition was set at 0, and tolerance for genome location was set as 10 (for more details please refer to section 2.5.2.1). Putative miRNA candidates were filtered during post-processing to increase robustness of the data, so that only miRNAs present in at least two of the three replicate EqPalF RNA libraries were retained for analysis.

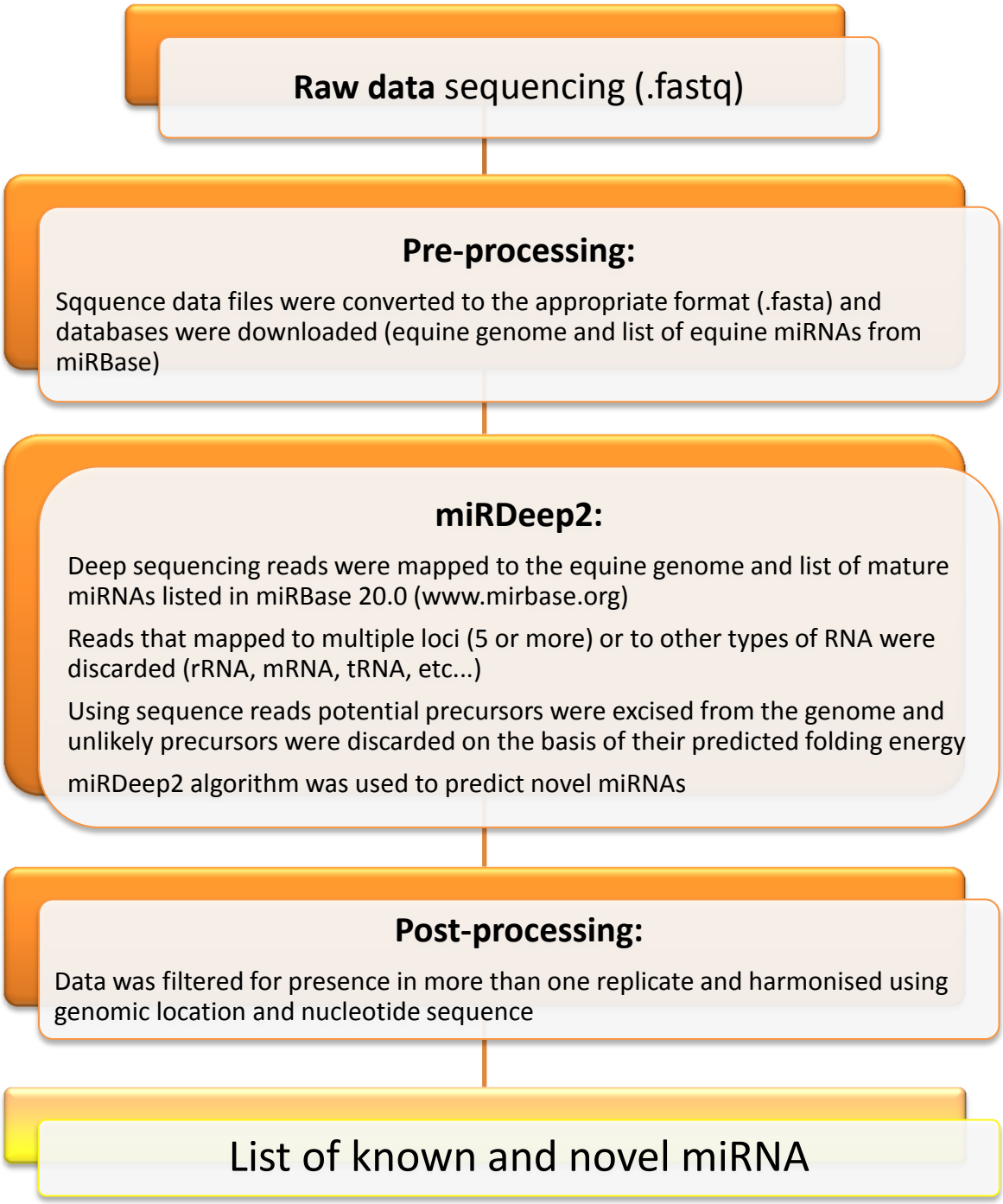


Figure 4.2. Summary of the steps of the analysis of the HTS data

As part of the descriptive analysis, the following parameters were calculated: length of mature miRNAs, percentage of GC nucleotides (GC content), chromosomal distribution and clustering, and relative abundance of the 3p or 5p arm of each miRNA. These were calculated using Microsoft Excel. When comparing the data to previous studies, t-test was applied and statistical significance was set at $p<0.05$. Statistical differences were calculated using

SPSS. Figures were generated using Excel, ClustVis (http://biit.cs.ut.ee/clustvis/) and the Genome Decorator Page online tool (NCBI).

To verify if the miRNAs were novel or had been described before, the list of miRNA candidates was analysed using two methods (BLAST+ and Edit distance) using the list of miRNAs from miRBase 20.0 and the list of equine miRNAs available in the literature. Full details of both methods are explained in Chapter 2, section 2.5.2.2. and a summary is displayed in Figure 4.3. Functional analysis was performed using Ingenuity Pathway Analysis (IPA) software with the list of non-novel miRNAs.

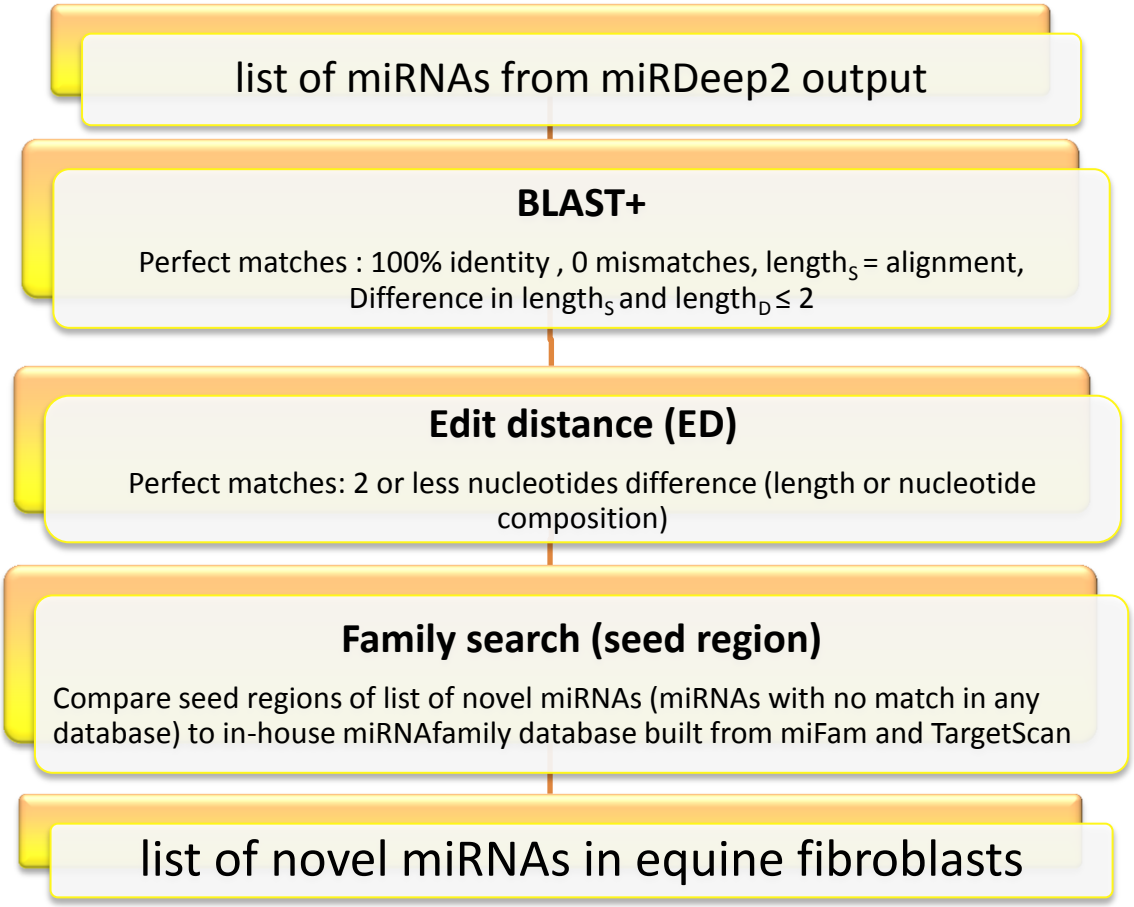


Figure 4.3. Summary of the pipeline for discovery of miRNAs in equine primary fibroblast (EqPalF)

4.3. Results

4.3.1. RNA quality, integrity and library quality control

Total RNA extracted from a flask of EqPalF cells maintained in culture showed good integrity as observed by electrophoresis on 1% agarose gel (Figure 4.4a). The A260/A280 ratio was 2.07 and the RIN value was 9.90, obtained using the Bioanalyser 2100 (Figure 4.4b) indicating suitable RNA for sequencing.

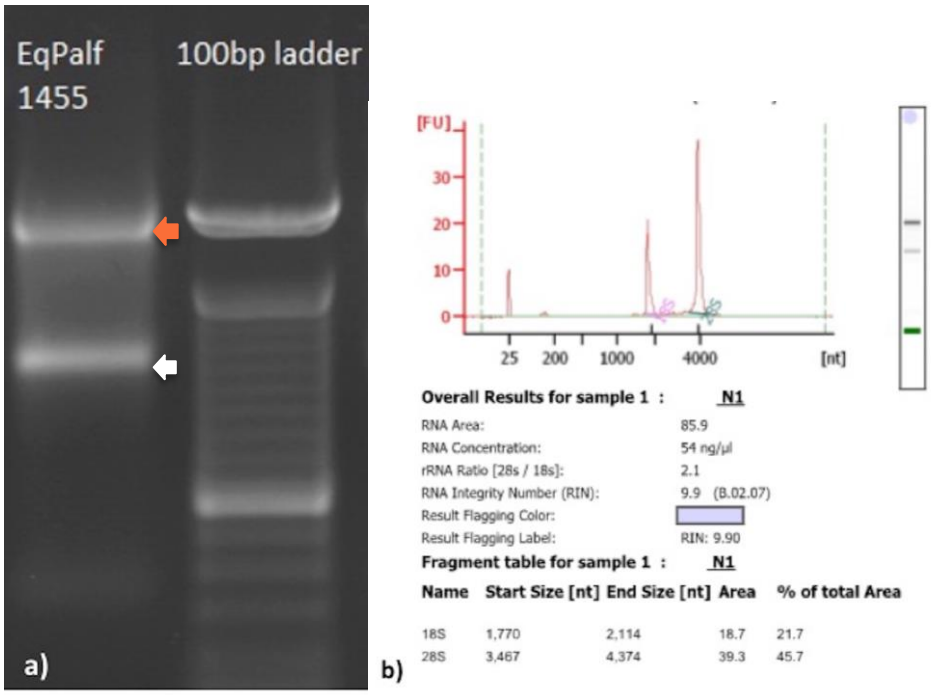


Figure 4.4. a) RNA sample from EqPalF visualised in a transilluminator on a 1% agarose gel. Good demarcation of the 28S (orange arrow) and 18S (white arrow) bands indicates high integrity of the RNA sample. **b) Electrophenogram obtained with Bioanalyzer 2100 of EqPalF RNA**. Single clear peaks correspond to the 18S and 28S fraction and no signs of RNA degradation are present as indicated by a high RIN value of 9.9.

The RNA extracted from the cultured EqPalF cells was divided into three samples for three technical replicates for RNA sequencing. During library preparation, two of the replicates, EqPalF.1 and EqPalF.3, demonstrated a unique peak of small RNAs in the DNA high specificity chip (band between 110 bp and 230 bp) (Figure 4.5a.). The second library, EqPalF.2 showed secondary peaks indicating possible contamination (Figure 4.5b). Preparation of this library was

subsequently repeated and the repeated DNA high sensitivity chip demonstrated a discrete small RNA peak on this second occasion (Figure 4.5c).

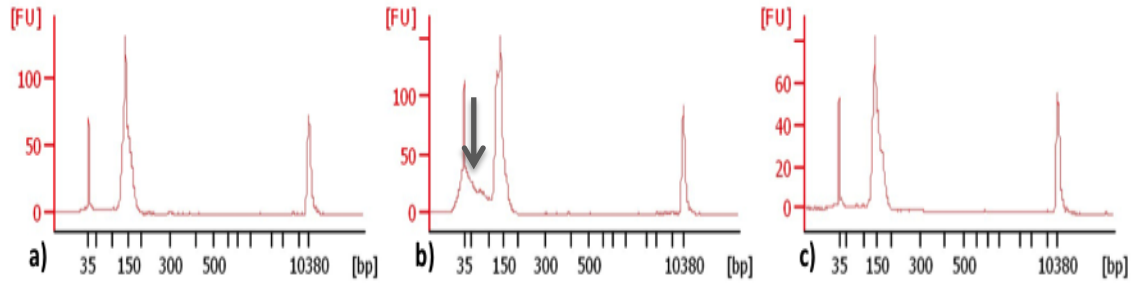


Figure 4.5. Image from the DNA high sensitivity chip showing the fraction of small RNA from the libraries. a) EqPalF.1 and EqPalF.3 both produced the same image on the DNA high sensitivity chip, showing a single peak around 150 bp signifying the fraction of small RNAs. b) EqPalF.2 had additional small peaks associated to the 150 bp peak which was compatible with contamination (grey arrow). c) EqPalF.2 library was repeated at a later time and the repeated DNA high sensitivity chip showed a single peak around 150 bp corresponding to the small RNA fraction with no contamination

4.3.2. Characterisation of RNA read data - correlation across replicate libraries, read count and number of mature miRNA candidates

The number of reads obtained from RNA sequencing ranged from 7.81 x 10^6 to 10.70 x 10^6 per library. After adaptor-trimming sequence reads were fed into miRDeep2. miRDeep2 mapped the sequencing reads to the databases (equine genome and equine miRNAs included in miRBase), removed redundancy and performed miRNA predictions based on a core algorithm and folding energy. Sequences that were not mapped to the equine genome, or mapped to other types of RNA such as tRNA, sequences that mapped to five or more loci in the genome, and sequences that did not meet the folding energy criteria set for miRNA in the Randfold script were removed. The number of reads retained after miRDeep2 and harmonisation ranged from 3.82 x 10^6 to 6.72 x 10^6 per library (see Table 4.2). Applying the filter of a candidate miRNA being present in at least two replicates of the cell line, a total of 593 unique sequences of putative mature miRNAs were identified in EqPalF RNA. Not all replicates expressed the 593 miRNA candidates. Library Replicate 2 (EqPalF.2) had a smaller number with 534 miRNAs in total, whereas Replicates 1 and 3 expressed 560 and 566 miRNAs,

respectively. The full list of miRNAs before and after data analysis is included in the file 'master file EqPalF.xlsx' in the DVD enclosed.

Table 4.2. Summary of reads and number of miRNAs per replicate library prepared from EqPalF cells

	Replicate 1 (EqPalF.1)	Replicate 2 (EqPalF.2)	Replicate 3 (EqPalF.3)
Number of reads from the sequencer	10,697,659	7,811,655	9,508,226
Number of reads after miRDeep2	6,716,925	3,820,310	6,125,119
Number of miRNA candidates from miRDeep2	638	583	628
Number of miRNA candidates after filtering for "at least present in two replicates"	560	534	566

The correlation between the number of reads for each miRNA between each replicate was calculated as part of quality control analysis of the data. The correlation between the three libraries was good as indicated by a Pearson's coefficient that ranged from 0.96-1.00 (see Table 4.3. and Figure 4.6.)

Pearson's coefficient is a measurement of linear correlation and the values range from -1.00 to +1.00. Two replicates with a correlation coefficient close to 1.00 indicates that the relative values of microRNA expression within each replicate are very similar, whereas closer to 0.00 would indicate that the two datasets are highly dissimilar.

Table 4.3. Pearson's correlation coefficient for EqPalF replicate libraries.
Coefficients were obtained by comparing the number of reads of each mature miRNA candidate in the three library replicates from EqPalF cells

	EqPalF.1	EqPalF.2	EqPalF.3
EqPalF.1	1	0.96	1
EqPalF.2	0.96	1	0.96
EqPalF.3	1	0.96	1

Read counts were obtained using the average value from the three replicates. Level of expression measured by read counts was variable from 1 read to 800,297 reads. Forty one per cent of the miRNA candidates had at least 100 reads. The 20 most highly expressed miRNAs in each replicate ranked by number of reads, are displayed in Table 4.4.The top 20 most highly expressed miRNAs in EqPalF using the average from the three replicates is presented in Table 4.5.

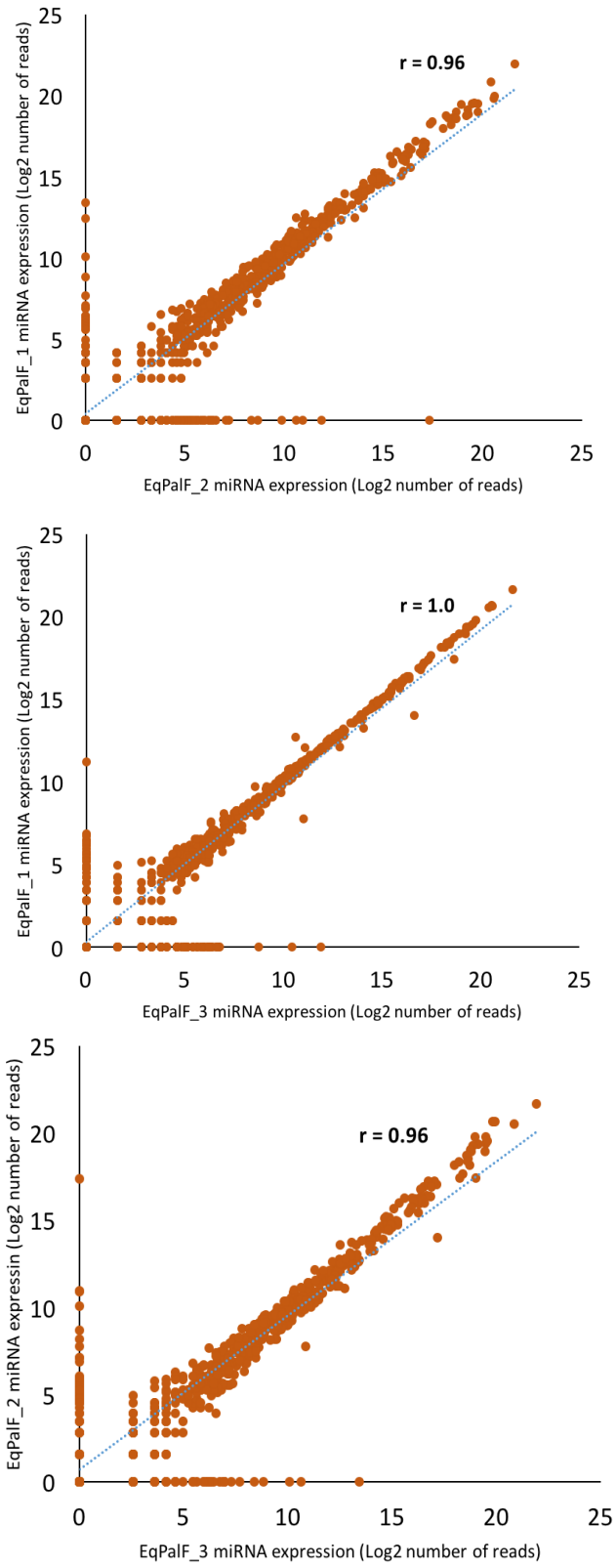


Figure 4.6. Correlation of miRNA expression in EqPalF replicates. The number of reads of each miRNA, expressed in $\log_2$ values from each of the three replicates was compared in a pair-wise manner. Each dot represents one miRNA with the $\log_2$ number of reads on the x axis from one replicate and on the y axis from a different replicate. Pearson's coefficient for each pair is displayed as 'r' and it shows that the linear relationship in the three pair comparisons is positive and strong (as values are close to +1.0)

Table 4.4. Top 20 highly expressed miRNAs in each replicate ranked by expression levels (number of reads)

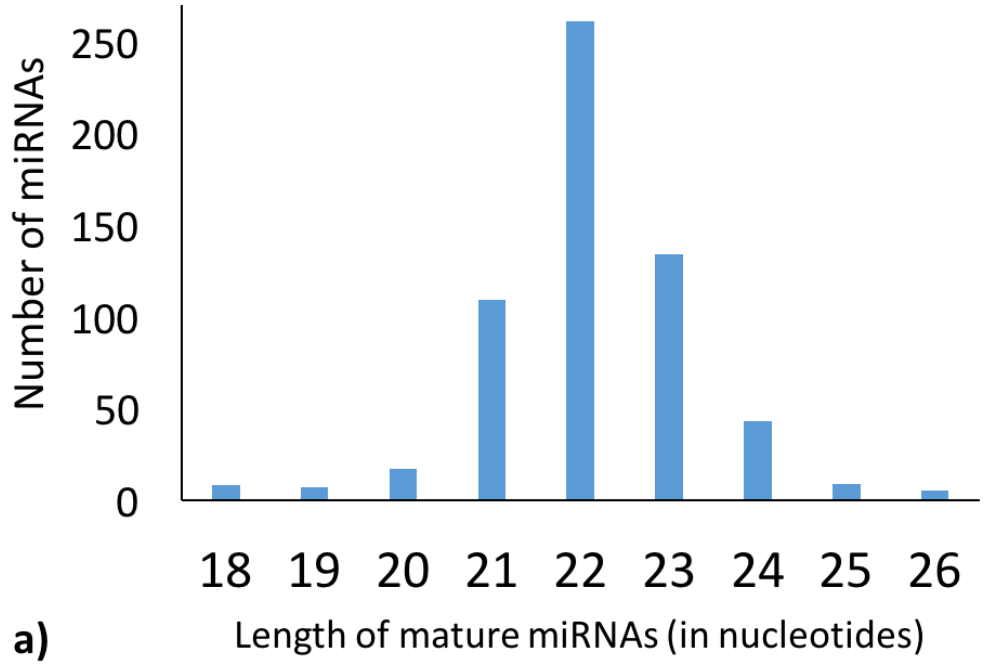
Internal ID	miRBase annotation	Expression in EqPalF.1	Internal ID	miRBase annotation	Expression in EqPalF.2	Internal ID	miRBase annotation	Expression in EqPalF.3
U0_3p	eca-miR-143	913,516	U0_3p	eca-miR-143	643,975	U0_3p	eca-miR-143	843,400
U4_5p	eca-miR-181a	445,006	U1_5p	eca-miR-21	300,211	U2_3p	eca-miR-22	423,803
U2_3p	eca-miR-22	431,534	U4_5p	eca-miR-181a	161,312	U4_5p	eca-miR-181a	420,779
U1_5p	eca-miR-21	391,725	U2_3p	eca-miR-22	150,179	U1_5p	eca-miR-21	385,487
U5_3p	eca-miR-27b	248,699	U9_3p	eca-miR-127	124,283	U5_3p	eca-miR-27b	227,871
U8_5p	eca-miR-26a	247,455	U5_3p	eca-miR-27b	117,996	U8_5p	eca-miR-26a	226,812
U9_3p	eca-miR-127	222,673	U10_3p	eca-miR-92a	117,017	U9_3p	eca-miR-127	197,266
U10_3p	eca-miR-92a	202,203	U3_5p	eca-miR-10b	115,856	U10_3p	eca-miR-92a	177,459
U12_3p	eca-miR-28-3p	179,825	U11_5p	eca-miR-191a	90,970	U11_5p	eca-miR-191a	172,656
U11_5p	eca-miR-191a	177,942	U15_5p	eca-miR-100	84,807	U12_3p	eca-miR-28-3p	161,028
U13_3p	eca-miR-222	173,475	U8_5p	eca-miR-26a	83,335	U14_5p	eca-miR-125a-5p	135,190
U14_5p	eca-miR-125a-5p	167,581	U12_3p	eca-miR-28-3p	77,051	U13_3p	eca-miR-222	130,282
U3_5p	eca-miR-10b	139,752	U13_3p	eca-miR-222	73,142	U3_5p	eca-miR-10b	130,248
U15_5p	eca-miR-100	116,157	U14_5p	eca-miR-125a-5p	71,908	U19_5p	eca-miR-30d	110,425
U19_5p	eca-miR-30d	116,048	U17_3p	eca-miR-221	70,097	U18_5p	eca-let-7a	95,835
U18_5p	eca-let-7a	102,475	U18_5p	eca-let-7a	65,586	U6_3p	eca-miR-199a-3p	88,258
U21_5p	eca-miR-125b-5p	98,581	U6_3p	eca-miR-199a-3p	64,069	U21_5p	eca-miR-125b-5p	84,015
U6_3p	eca-miR-199a-3p	89,846	U19_5p	eca-miR-30d	62,668	U17_3p	eca-miR-221	75,235
U17_3p	eca-miR-221	83,314	U22_5p	eca-miR-145	54,596	U16_5p	eca-miR-16	74,513
U16_5p	eca-miR-16	73,835	U24_3p	eca-miR-23b	49,473	U22_5p	eca-miR-145	52,295

Table 4.5. Top 20 highly expressed miRNAs in EqPalF, ranked by average of expression in three replicates

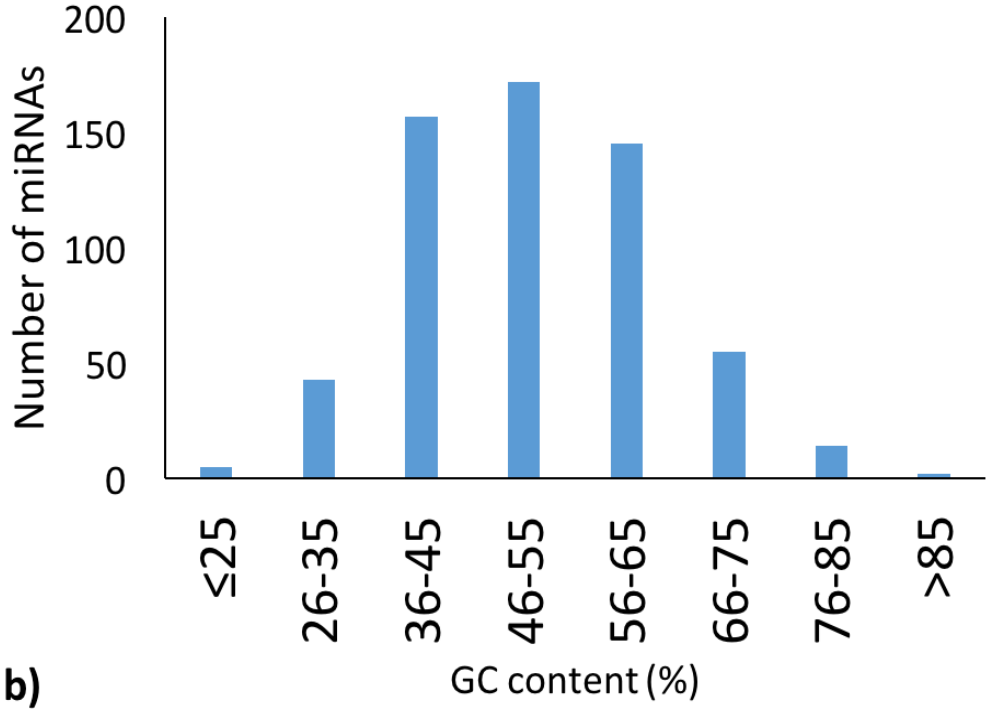
Internal ID	Mature sequence	Annotation (BLAST-ED method)	Average number of reads
U0_3p	UGAGAUGAAGCACUGUAGCUC	eca-miR-143	800,297
U1_5p	UAGCUUAUCAGACUGAUGUUGA	eca-miR-21	359,141
U4_5p	AACAUUCAACGCUGUCGGUGAGU	eca-miR-181a	342,366
U2_3p	AAGCUGCCAGUUGAAGAACUGU	eca-miR-22	335,172
U5_3p	UUCACAGUGGCUAAGUUCUGC	eca-miR-27b	198,189
U8_5p	UUCAAGUAAUCCAGGAUAGGCU	eca-miR-26a	185,867
U9_3p	UCGGAUCCGUCUGAGCUUGGCU	eca-miR-127	181,407
U10_3p	UAUUGCACUUGUCCCGGCCUGU	eca-miR-92a	165,560
U11_5p	CAACGGAAUCCCAAAAGCAGCUG	eca-miR-191a	147,189
U12_3p	CACUAGAUUGUGAGCUCCUGGAG	eca-miR-28-3p	139,301
U3_5p	UACCCUGUAGAACCGAAUUUGU	eca-miR-10b	128,619
U13_3p	AGCUACAUCUGGCUACUGGGUCUC	eca-miR-222	125,633
U14_5p	UCCCUGAGACCCUUUAACCUGUGA	eca-miR-125a-5p	124,893
U19_5p	UGUAAACAUCCCCGACUGGAAGCU	eca-miR-30d	96,380
U18_5p	UGAGGUAGUAGGUUGUAUAGUU	eca-let-7a	87,965
U15_5p	AACCCGUAGAUCCGAACUUGUG	eca-miR-100	82,070
U6_3p	ACAGUAGUCUGCACAUUGGUU	eca-miR-199a-3p	80,724
U21_5p	UCCCUGAGACCCUAACUUGUGA	eca-miR-125b-5p	76,852
U17_3p	AGCUACAUUGUCUGCUGGGUUUC	eca-miR-221	76,215
U16_5p	UAGCAGCACGUAAAUAUUGGCG	eca-miR-16	63,372

4.3.3. Mature sequence length and GC content

The sequences of the putative miRNAs in EqPalF ranged between 18 and 26 nucleotides in length and the median length was 22 nucleotides (see Figure 4.7.a). The percentage of GC nucleotides in the putative miRNA sequences varied, ranging from 19.05% -100.00% (average 51.63%, median 50%, Figure 4.7.b). The length distribution of mature miRNA candidates in this study was consistent with the distribution of equine miRNAs in miRBase 20.0 (*in silico* predicted and HTS) (Zhou *et al.*, 2009; Platt *et al.*, 2014) as well as experimental data from other species (Ji *et al.*, 2012; Shukla *et al.*, 2015). With the data available from the two previous HTS equine studies it is not feasible to compare miRNA by miRNA but overall, the length distribution was also similar to that described for miRNAs in equine bone and cartilage (Desjardin *et al.*, 2014) and to colon, liver and muscle in the horse (Kim *et al.*, 2014) when including both known miRNAs and 'novel miRNAs' reported in both studies. GC content was consistent with previous studies in equine miRNAs (Zhou *et al.*, 2009; Desjardin *et al.*, 2014; Kim *et al.*, 2014).



a)



b)

Figure 4.7. Length distribution of miRNAs and GC nucleotide content (%) of 593 miRNAs found in EqPalF cells

4.3.4. Distribution of microRNAs within the equine genome and distribution of 5p/3p miRNA arms

The equine genome comprises 32 chromosomes (http://www.ncbi.nlm.nih.gov/genome?term=equus%20caballus). All chromosomes but one (ECA 31) encoded putative microRNAs. Chromosomes 24, X, 11 and 7 had the highest number of microRNAs with 99, 64, 47 and 41 entries respectively, whereas chromosomes 18, 21, 27 and 29 had the lowest number of miRNA candidates, seven, four, four and one, respectively. No miRNAs were found on chromosome 31. A histogram showing the number of miRNAs encoded on each chromosome is shown in Figure 4.8.

The majority of the mature miRNA candidates mapped to a single genomic locus (545 miRNAs, 91.95%), whereas a small proportion were found in two (40 miRNAs, 6.75%), three (6 miRNAs, 1.01%) and four loci (2 miRNAs, 0.34%). As stated above, miRNAs mapping to five or more loci were removed by the miRDeep2 software. The chromosomal distribution of miRNAs in EqPalFs was in accordance with the equine miRNAs included in the miRBase (Zhou *et al.*, 2009; Platt *et al.*, 2014) and similar to previous studies undertaken in colon, liver and muscle (Kim *et al.*, 2014) and bone and cartilage in horses (Desjardin *et al.*, 2014) (Figure 4.9). A notable exception is the abundance of miRNAs identified on chromosome 24 in the current study. 15.21% of total miRNAs (99 miRNA candidates) from EqPalF RNA sequencing were identified on chromosome 24, which far exceeds that in previous equine miRNA studies.

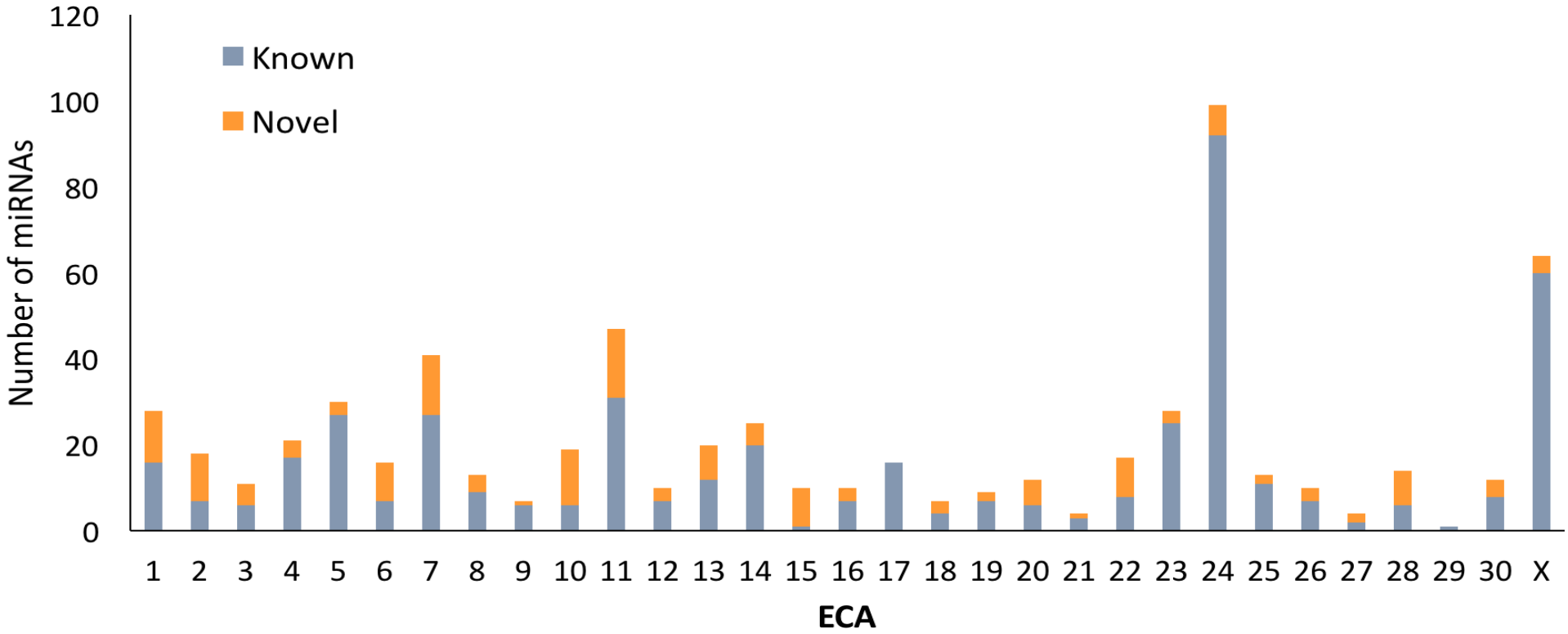


Figure 4.8. Histogram representing 593 miRNA candidates mapped to the 32 chromosome in the horse. miRNAs that were found in more than one locus have been included in all locations that they were mapped to (593 miRNAs, 636 loci). miRNAs which mapped to the non-characterised areas of the genome ("Un", for more details see horse genome assembly EqCab2.0 in http://www.ncbi.nlm.nih.gov/genome?term=equus%20caballus) were omitted (15 miRNAs)

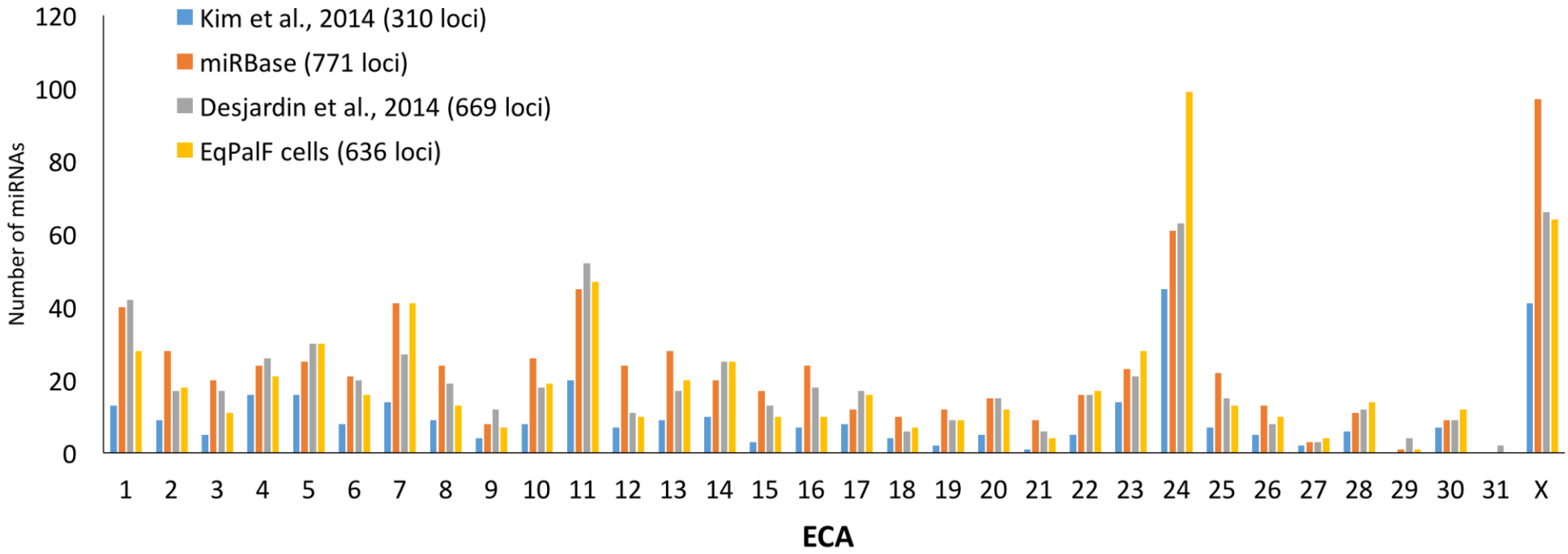


Figure 4.9. Histogram representing the comparison of chromosomal distribution of 593 miRNA candidates sequenced from EqPalF cells mapped to 636 loci to miRBase 20.0 and two equine HTS studies. The average distribution from three EqPalF replicates was compared to equine miRNAs from the miRBase 20.0 (690 mature sequences, 771 loci), and from two HTS studies on equine miRNAs (Desjardin et al. (2014) with 669 mature sequences, 371 known and 298 suggested as novel, mapped to 669 loci; and Kim et al. (2014) with 292 mature sequences mapped to 310 loci as no information is provided about the genomic location of the suggested novel miRNAs in the latest study

Interestingly, several miRNAs were found to be located in close proximity to one another, suggesting that these may be encoded as polycistronic miRNAs. This can be appreciated by the close chromosomal positioning of some miRNAs shown in the ideogram in Figure 4.10. A total of 255 miRNAs (43%) were found forming 43 clusters. Clusters were defined as groups of miRNAs encoded in the same strand at not more than 3 Kb apart (Zhou *et al.*, 2009; Kim *et al.*, 2014). The distribution of miRNA clusters was found to be similar to previous studies although in the current study no clusters were found on ECA 1 (Zhou *et al.*, 2009), and ECAs 16, 18, 20 and 28 (Kim *et al.*, 2014) (see Figure 4.11). ECAs 24 and X had the highest density of miRNA clusters identified from EqPalF cells and in particular two main areas of ECA 24 with high density of miRNAs were identified and are shown by intense red bands in Figure 4.10.

On multiple occasions, miRNAs identified by sequencing mapped to the same genomic location and were encoded by either the 3' or 5' arm of the precursor miRNA. In total, 302 miRNAs mapped to the 3p end of the precursors (50.93%) and 291 miRNAs mapped to the 5p end of the precursors (49.07%).

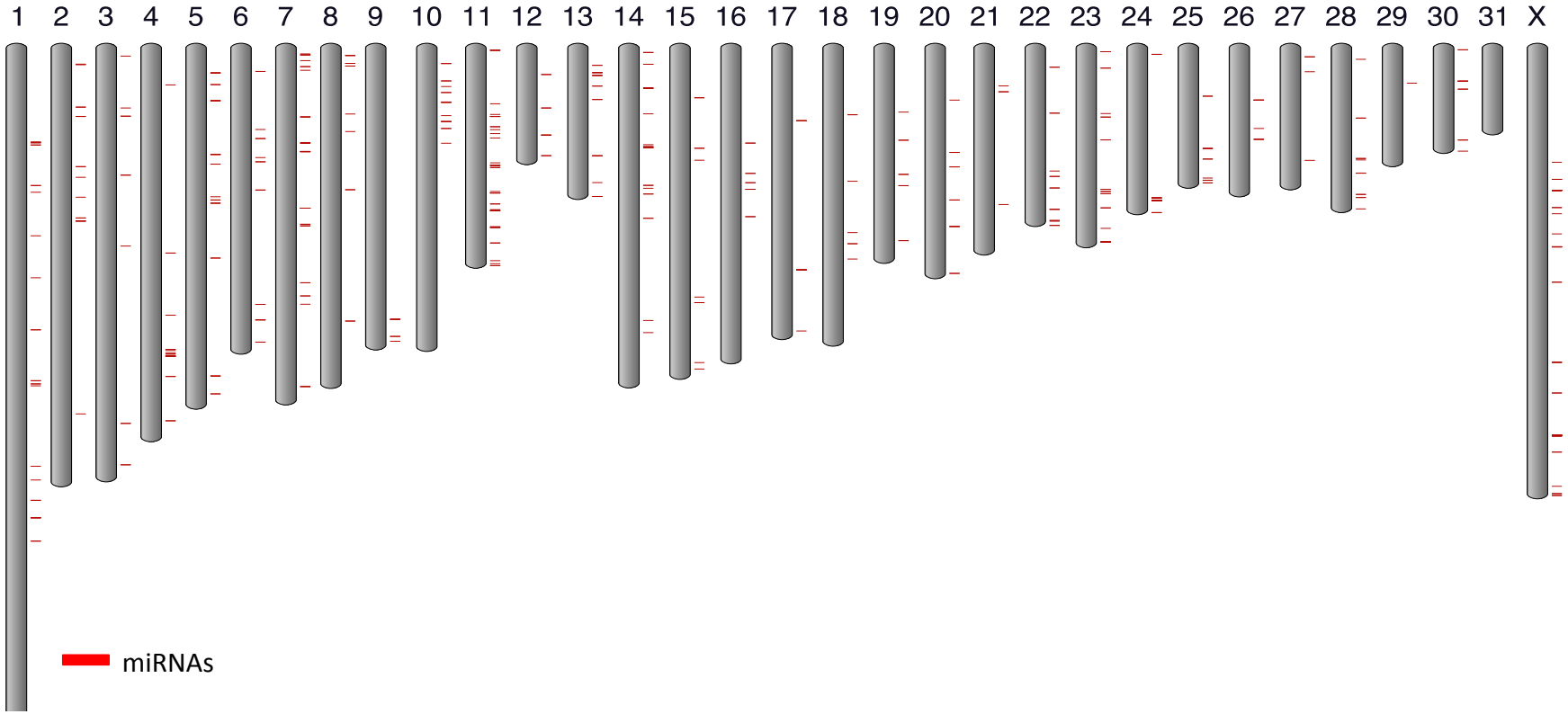


Figure 4.10. Ideogram showing the distribution and density of miRNA candidates in EqPalF cells when mapped to the 32 equine chromosomes. The broader the red band, the higher number of miRNAs that are encoded in the region. In total 255 were found in 43 clusters (<3Kb genomic distance). The chromosome with highest number of miRNAs (ECA24) had a high density of miRNAs, encoded in three areas of the chromosome. miRNAs which genomic location mapped to the non-characterised areas of the genome of the horse (named as "Un", for more details see horse genome assembly EqCab2.0. in http://www.ncbi.nlm.nih.gov/genome?term=equus%20caballus) were omitted (15 miRNAs)

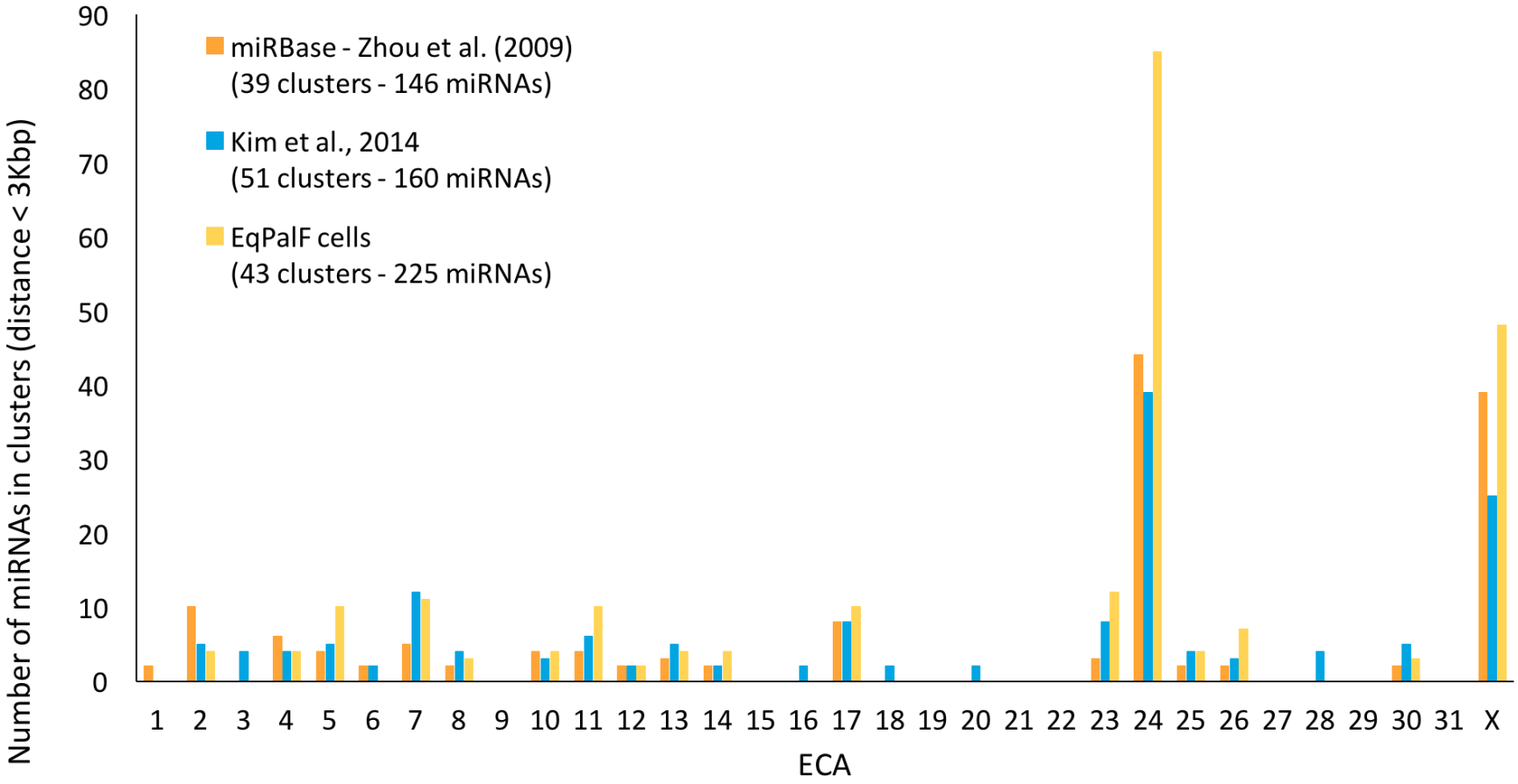


Figure 4.11. Histogram representing miRNA candidates from EqPalF cells mapped in clusters compared to those from Zhou *et al*. (2009) and Kim *et al*. (2014)

4.3.5. Characterisation of miRNAs in EqPalF cells: expression levels and novel miRNAs

Using a sequential method (first BLAST+ and then Edit distance), out of 593 miRNAs sequenced from EqPalF cells and mapped to the horse genome, 258 sequences were found to be already present in miRBase 20.0 as equine miRNAs (eca-miRs). Of the remaining 335 miRNAs sequenced, 71 matched human annotations, 18 matched mouse annotations, 73 matched other species and 173 had no match in any database and therefore are considered novel miRNAs. A small proportion of the novel miRNAs (22) could be grouped into miRNA families by aligning the seed region (nucleotides 2-7)(see Chapter 2). The full list of 593 miRNAs expressed in EqPalF cells is supplied in this thesis as a digital copy under the file of EqPalF_HTS_data.xlsx.

miRNAs were ranked by level of expression (average number of reads from the three replicate libraries). The top 20 most expressed miRNAs showing their miRBase annotation and level of expression (in number of reads, average from three replicates) is displayed in Figure 4.12. This histogram highlights the abundance of eca-miR-143 which represents 17% of the total number of reads from EqPalF cells. No novel miRNAs were amongst the top 20 most expressed miRNAs.

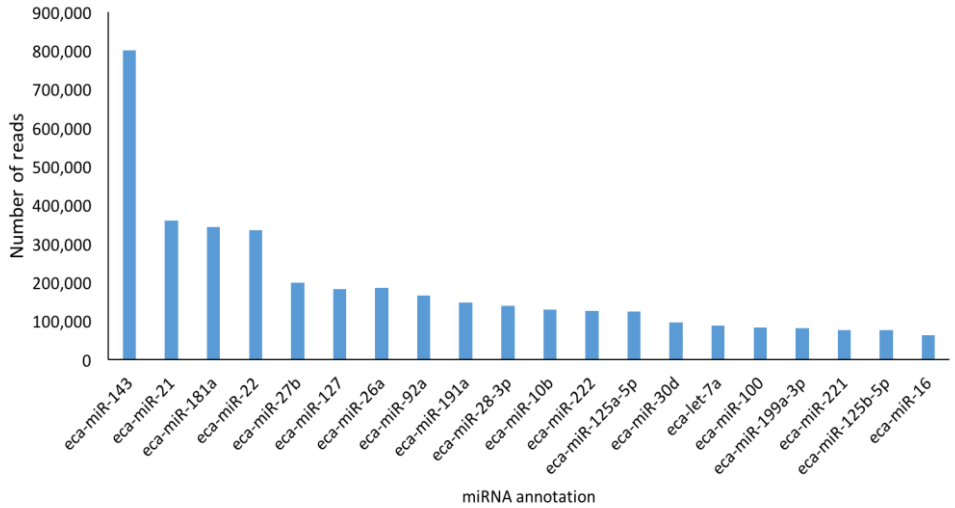


Figure 4.12. Histogram representing the top 20 most expressed miRNAs in equine primary fibroblasts (EqPalF cells) ranked by number of reads. Results are expressed as the average of reads from three replicate libraries

Out of 593 miRNA candidates obtained in the current study, 334 had been previously assessed using microarray (please see Chapter 3 for more details). Of those, 143 were also found to be present in EqPalF cells during the microarray analysis. The remaining 191 were indeed interrogated during the microarray although they produced signals below the background level. Over half of those miRNAs (58%, 109 miRNAs) had low expression (≤100 reads) on the HTS data, whereas over one third (36%, 69 miRNAs) had medium expression (101-1000 reads) and the remaining 6% (13 miRNAs) were expressed at least in 1000 reads (13 miRNAs).

Amongst the top 50 most expressed miRNAs there were three miRNAs that had never been reported in the horse, but previously described in other species: ssc-miR-151-3p, bta-miR-6119-5p and cgr-miR-140-3p with average read counts of 33361, 26381 and 12304, respectively. Notably, the level of expression of the novel miRNAs was low and ranged from 1 read to 3576 reads. The top 12 most expressed novel miRNAs ranged from 100 to 3576 reads.

4.3.6. Functional analysis

The list of 593 miRNAs sequenced from EqPalF cells were transformed to their human, mouse or other species counterparts where possible (up to a maximal difference of 2 nucleotides either in length or in sequence composition) and uploaded onto Ingenuity Pathway Analysis software (IPA) for functional analysis. This software does not list equine miRNA per se and is based on a generic mammalian database, thus this adaptation was necessary to facilitate functional analysis. In total, 256 miRNAs were present in the IPA database. The core analysis function, which tests for enrichment of a gene-set within the various forms of database contained in IPA provided no usable results and therefore pathways of interest and upstream regulators could not be identified. The Target filter analysis tool recognised 149 miRNAs, which can potentially target 15967 targets. The top diseases and functions highlighted by this analysis were ranked by score and number of molecules identified: cancer of the reproductive system, immunological disease of muscular disorders, cellular development, proliferation and growth, cancer of the gastrointestinal system, cell cycle dysregulation. Results are displayed in Table 4.6.

Table 4.6. List of networks that were highlighted in EqPalF using IPA software. Data obtained from using as input 593 miRNAs, of which 256 were listed by the software as "mapped" and 149 had allocated targets. Core analysis includes not only miRNAs present in the data set but other genes/miRNA targets that are directly or indirectly linked to those miRNAs

Score	Focus Molecules	Top Diseases and Functions
55	28	Cancer, Organismal Injury and Abnormalities, Reproductive System Disease
36	21	Organismal Injury and Abnormalities, Reproductive System Disease, Cancer
29	16	Cell Death and Survival, Skeletal and Muscular Disorders, Immunological Disease
23	16	Organismal Injury and Abnormalities, Reproductive System Disease, Cancer
21	14	Cellular Development, Cellular Growth and Proliferation, Cell Death and Survival
21	14	Organismal Injury and Abnormalities, Reproductive System Disease, Developmental Disorder
21	14	Cancer, Gastrointestinal Disease, Organismal Injury and Abnormalities
20	14	Cellular Development, Cellular Growth and Proliferation, Cancer
18	13	Cancer, Organismal Injury and Abnormalities, Cell Cycle
18	13	Developmental Disorder, Hereditary Disorder, Organismal Injury and Abnormalities
17	12	Cancer, Organismal Injury and Abnormalities, Reproductive System Disease
17	13	Cancer, Organismal Injury and Abnormalities, Reproductive System Disease
15	11	Cancer, Organismal Injury and Abnormalities, Cellular Development
4	2	Neurological Disease, Psychological Disorders
4	2	Cancer, Gastrointestinal Disease, Hepatic System Disease
4	2	Cardiovascular Disease, Congenital Heart Anomaly, DNA Replication, Recombination, and Repair
2	1	Endocrine System Disorders, Organismal Injury and Abnormalities, Reproductive System Disease
2	1	Cancer, Gastrointestinal Disease, Organismal Injury and Abnormalities
2	1	Cancer, Organismal Injury and Abnormalities, Hematological Disease
2	1	Cancer, Organismal Injury and Abnormalities, Reproductive System Disease
2	1	Connective Tissue Disorders, Inflammatory Disease, Inflammatory Response
2	1	Cellular Response to Therapeutics, Connective Tissue Disorders, Inflammatory Disease
2	1	Dermatological Diseases and Conditions, Hereditary Disorder, Organismal Injury and Abnormalities

4.4. Discussion

To begin to understand the roles of miRNAs in equine sarcoids, the miRNA transcriptome of equine primary fibroblasts was characterised. In the current chapter, expression and abundance of 593 putative miRNA candidates in EqPalF cells were described using the bioinformatic analysis of HTS data. Of those, 258 had previously been described from the horse, with sequences available in miRBase 20.0. Amongst the remaining 335 miRNA candidates, 162 were homologous to miRNAs identified in other species and 173 have not been reported before in the horse or other species. All novel miRNAs made a stable hairpin structure as predicted by miRDeep2 using the prediction algorithm. Therefore this represents the first report of novel miRNAs found in equine primary cells using HTS.

Currently HTS is used as a tool for the identification and discovery of miRNAs in various organisms as it has several advantages compared to previous methods including high throughput, high repeatability, low background noise, ability to detect miRNAs expressed at low levels, and identification of previously uncharacterised miRNA sequences (Morozova & Marra, 2008; Yan *et al.*, 2011; Chen *et al.*, 2012; Gunaratne *et al.*, 2012). Successful RNA extraction and library construction are two of the most crucial steps required for a suitable set of sequencing data (Metzker, 2010; Pritchard *et al.*, 2012). In the current study the integrity and quality of RNA prepared from equine primary fibroblasts was good. During library construction one of the replicates, EqPalF.2 had to be repeated at a later time due to suspected contamination. This led to small differences in the sequencing data obtained as seen in the marginally lower correlation coefficients (0.96) and marginally lower number of miRNAs identified in this replicate (534 *vs.* 560 and 566 for EqPalF.2 *vs.* EqPalF.1 and EqPalF.3 respectively). However, despite differences being observed between the other two replicates and EqPalF.2, the correlation between all three replicates was excellent (0.96 – 1.0), indicating that the results are robust.

The length distribution of mature miRNA candidates (18-26 nucleotides) and the GC nucleotide content were in agreement with previously predicted equine miRNAs in miRBase and very similar to that previously predicted and reported by

HTS from equine bone, cartilage, colon, liver and muscle (Zhou *et al.*, 2009; Desjardin *et al.*, 2014; Kim *et al.*, 2014) and to miRNAs in other species (Ji *et al.*, 2012; Shukla *et al.*, 2015). This comparison was made with the full list of miRNAs provided by the publishers that include 'known miRNAs' as well as 'novel miRNAs'. It needs to be stated that a considerable number of miRNAs from Desjardin *et al.* (2014) (298 mature miRNA sequences) and Kim *et al.* (2014) (329 mature miRNA sequences) were not included in the official database at the time the analysis was performed but were nevertheless included in my calculations for length and nucleotide composition comparisons. Dissimilarities in length, nucleotide composition and precise nomenclature were found between these studies and the official database. These differences restricted some of the comparisons during the data analysis, as by definition a difference in one or more nucleotides between sequences can result in them being considered as distinct miRNAs (Griffiths-Jones *et al.*, 2006).

The chromosomal distribution of the miRNAs expressed in equine primary fibroblasts was very similar to that reported in equine tissues and predicted in the equine genome (Zhou *et al.*, 2009; Desjardin *et al.*, 2014; Kim *et al.*, 2014). However, the results from these comparisons are to be taken with caution due to the same reasons as stated above when comparing length and nucleotide sequence. Additionally, one of the studies, Kim et al., (2014), only provided details of chromosomal coordinates for the known miRNAs (292 miRNAs encoded in 310 loci).

An interesting finding from examination of the chromosomal distribution of miRNAs in EqPalFs was that 48 miRNA candidates (6.75% of the total number of miRNAs) mapped to multiple locations. One of the assumptions made when processing the data using miRDeep2 is that sequences found five or more times along the genome are discarded and therefore the maximum number of loci that a miRNA could be encoded was limited to a maximum of four. From a computational perspective, these are considered repetitive DNA sequences rather than real miRNAs and excluded from the analysis.

Although the average expression for the miRNAs encoded by multiple loci was higher than the average expression of single locus miRNAs (7150 *vs.* 1593), the difference was not statistically significant. It can be hypothesised that some of

the miRNAs that are encoded in multiple parts of the equine genome regulate crucial pathways and therefore the multi-copy nature of the loci encoding these miRNAs may be viewed as a defence mechanism whereby if any damage or deletions occur to those loci, there is the possibility for the cellular machinery to switch the expression to alternative loci.

miRNAs are frequently found in clusters and these clustered miRNAs are believed to share a common primary transcript and have hence been suggested to be co-regulated and co-expressed (Altuvia *et al.*, 2005; Baskerville & Bartel, 2005). Clusters that contain miRNAs from the same family, as they share the same seed region, have been suggested to be gene duplication events with the purpose of regulating the same mRNA function (Maher *et al.*, 2006; Berezikov, 2011; Marco *et al.*, 2013). Interestingly miRNA clusters can contain miRNAs from different families and this is thought to be again a duplication event but with a slightly different purpose: instead of targeting the same mRNA, clustered miRNAs from different families may regulate a pathway by targeting multiple molecules that intervene in the pathway (Berezikov, 2011). It is thought that miRNAs encoded in distances shorter than 50 Kb derive from the same transcript and therefore may be co-expressed (Baskerville & Bartel, 2005). miRNA clusters are defined as miRNAs that map to adjacent chromosomal loci within a distance that ranges from approximately 1 Kbp to 10 kbp (Altuvia *et al.*, 2005; Zhou *et al.*, 2009; Kim *et al.*, 2014; Shukla *et al.*, 2015). For the analysis of clusters expressed in EqPalF cells, I used the criteria from Zhou *et al.* (2009) and Kim *et al.* (2014) as there are the only two studies that evaluated miRNA clusters in the equine genome and the sequences from Zhou et al. (2009) are present in the miRBase database. Using this criteria, 255 miRNA candidates from EqPalF cells were found to form 43 clusters. Unfortunately there was not a sufficient number of clustered miRNAs recognised by the IPA software to perform functional analysis. Further studies of equine miRNA clusters transcribed in the same transcripts will be required to evaluate target identification and co-regulation of functional pathways.

One of the most challenging aspects of the data analysis was to annotate the mature miRNA candidates. Nucleotide composition and length were utilised to ascertain whether the miRNA candidates found in EqPalFs had been reported

before in the horse or in any other species. Due to the quality of data available from the two studies that specifically assessed miRNAs in equine tissues using high throughput sequencing (HTS) (Desjardin *et al.*, 2014; Kim *et al.*, 2014), the results from the comparisons using nucleotide composition and length were inconsistent. When allocating a name/annotation to each miRNA candidate found in EqPalF cells, if a match was found in either the official database miRBase 20.0 (Zhou *et al.*, 2009; Platt *et al.*, 2014) or previous HTS studies of equine miRNAs (Desjardin *et al.*, 2014; Kim *et al.*, 2014), the annotation present in the official database was considered the most accurate. There is no consensus on what is sensible or acceptable when determining if two relatively similar sequences are indeed two distinct miRNAs. Two miRNA with differences in the nucleotide composition and/or length will have variation in the seed region and therefore their function may be affected (Berezikov, 2011). In the current study, I used a maximum deviation of two nucleotides in length for the BLAST+ method and a maximum deviation of two nucleotides in either length or composition for the ED method. This cut-off point was set after assessing the differences between the dataset obtained from equine primary fibroblasts, the official miRNA database and the list of miRNAs from two previous equine HTS studies (Desjardin *et al.*, 2014; Kim *et al.*, 2014) and after learning that commercial sequencing services use two as the cut-off to define the difference between two distinct miRNAs. This approach has the limitation that isomiRs may be mistakenly assigned different annotations, as additions or deletions to the 3' and 5' ends may simply be a variation during the biogenesis of a particular miRNA and unequivocally this will have an effect on the miRNA function (Manzano *et al.*, 2015).

In total, 143 miRNAs expressed by EqPalF cells were identified by both HTS and microarray methods (for details on microarray results, the reader is referred to Chapter 3). However, out of the 334 sequences that were common to the two assays (identified by HTS / oligos included in the microarray chip design), 191 were found in the sequencing data but did not produce a signal above background noise on the microarray. More than half of those were expressed at a low level in EqPalF cells in the HTS dataset (58%, 109 miRNAs, ≤100 reads). The probes in the microarray were designed with miRNA sequences from the miRBase from the horse (predicted miRNAs) and from the human datasets. Of the miRNAs

expressed at >1000 reads in HTS but not detected during microarray (6%, 13 miRNAs), it is likely that mismatches in the nucleotide composition between the miRNA sequence in EqPalF and the microarray probe could explain the presence of those miRNAs in the HTS set and low or undetectable levels in the microarray. Moreover, HTS methods can detect miRNAs present at a low level that may be overlooked using other techniques. Additionally, 259 miRNAs were sequenced that were not included in the microarray design. Again, this highlights another advantage of the HTS over microarray, as it does not rely on previous knowledge of miRNA sequences.

Amongst the expressed miRNAs in equine primary fibroblasts, the miRNAs provisionally named as eca-miR-21, eca-miR-22, eca-miR-26a, eca-miR-27b, eca-miR-28-3p, eca-miR-92a, eca-miR-127, eca-miR-143, eca-miR-181a and eca-miR-191a were the 10 most abundantly expressed miRNA candidates. Bioinformatic analysis identified five of these miRNAs to be encoded in clusters in the equine genome: cluster miR-143/145, cluster miR-181f/a/b, cluster miR-24/27b/23b, cluster miR-433/127/432/136 and cluster miR-17/106/18/19/20/92. miRNAs encoded by these clusters have been reported to be expressed in equine colon, liver and muscle tissue (Zhou *et al.*, 2009; Kim *et al.*, 2014). Therefore, the current study represents the first experimental verification of the presence of those miRNAs in equine primary fibroblasts and contributes to the knowledge of miRNA clustering in the horse.

The most highly expressed miRNA, eca-mir-143, was found to represent 17.2% of the total reads. Significant upregulation of miR-143 has been reported in fibroblasts and smooth muscle cells in a colorectal cancer model when compared to other types of cells using HTS technology (Kent *et al.*, 2014). Interestingly, the orthologue of this miRNA in the goat has been found to be the most abundant miRNA in the mammary gland (Ji *et al.*, 2012). Downregulation of miR-143/145 has been reported in colorectal cancer, breast cancer and cervical cancer, and an anti-tumoural role has been attributed to these miRNAs through regulation of MAPK and p53 signalling pathways and hexokinase 2 (Suzuki *et al.*, 2009; Fang *et al.*, 2012; Liu *et al.*, 2012; Pagliuca *et al.*, 2013; Das & Pillai, 2015). Further investigations will be required to ascertain the roles of miR-143 in equine primary fibroblasts.

In summary, the results presented in this chapter reveal the expression profile of equine miRNAs in primary fibroblast cells. This provides experimental verification of the existence of 258 equine miRNAs and the presence of an additional 335 novel miRNAs in the horse. Out of the 335, 173 of the miRNAs identified here were not only novel for the horse but for any other species, as no annotation match is available in miRBase. Having established the expression profile of miRNAs in normal equine fibroblast cells, this information can be applied to examine any changes in miRNA expression in BPV-1 transformed cells and sarcoid cell lines. This can be achieved using HTS and qRT-PCR, as detailed in the next chapter.

Chapter 5. Differential expression of miRNAs in equine sarcoids: high-throughput sequencing of the miRNAs in equine primary fibroblasts, equine sarcoid fibroblasts and equine fibroblasts transformed *in vitro* with Bovine Papillomavirus type 1 genome

5.1. Introduction

Equine sarcoids (ES) are a recognised health and welfare concern in horses and other *Equidae* and represent the most common cutaneous tumour in horses, donkeys, and mules (van Dyk *et al.*, 2009; Wobeser *et al.*, 2010; Schaffer *et al.*, 2013; Sanchez-Casanova *et al.*, 2014; NEHS, 2015). The aetiology is intrinsically linked to bovine papillomavirus (BPV) types 1 and 2 (Nasir & Reid, 1999; Yuan *et al.*, 2007a; Brandt *et al.*, 2011b; Wobeser *et al.*, 2012). Risk factors such as age, breed, equine leucocyte antigen haplotype and other genetic influences (single nucleotide polymorphisms and DNA deletions and duplications) have also been identified (Meredith *et al.*, 1986; Brostrom *et al.*, 1988; Mohammed *et al.*, 1992; Brostrom, 1995; Wobeser *et al.*, 2010; Jandova *et al.*, 2012; Bugno-Poniewierska *et al.*, 2016; Staiger *et al.*, 2016). An in-depth review of the current knowledge of equine sarcoids was presented in Chapter 1.

BPV types associated with ES belong to a family of heterogeneous viruses that can give a great variety of clinical presentations from benign lesions to malignant cancer (Silvestre *et al.*, 2009; Pangty *et al.*, 2010; Maiolino *et al.*, 2013; Roperto *et al.*, 2013). The oncogenic effects of papillomaviruses have been extensively studied and recently reviewed for HPV and BPV (Moody & Laimins, 2010; Venuti *et al.*, 2011; DiMaio & Petti, 2013; Banno *et al.*, 2014; de Freitas *et al.*, 2014; Gil da Costa & Medeiros, 2014; Munday, 2014; Doorbar *et al.*, 2015; Bocaneti *et al.*, 2016). The oncogenic activity in high-risk HPVs -16 and -18 (HR-HPVs) is attributed to oncoproteins E7, E6 and to a lesser extent, E5, whereas in BPV-1/-2 it is primarily attributed to E5 and E6 and to a lesser extent, E7 (Moody & Laimins, 2010; Yuan *et al.*, 2011b). The best-studied oncogenic mechanisms for HR-HPVs are the downregulation of p53 tumour suppressor protein and pRB pathway by E6 and E7 respectively, whereas for BPV-1/-2 it is the activation of PDGFß-R by BPV E5 (Borzacchiello *et al.*, 2009; Marchetti *et al.*, 2009; Moody & Laimins, 2010; Altamura *et al.*, 2013). Both HR-HPVs and BPV-1/-2 prevent expression of MHC class I on the cell surface, promote activation of the MAPK-p38 cascade, affect expression of p53 and stimulate anchorage-independent growth (Ashrafi *et al.*, 2002; Yuan *et al.*, 2008a; Venuti *et al.*, 2011; Yuan *et al.*, 2011c; Altamura *et al.*, 2013; de Freitas *et al.*, 2014). Interestingly, some of the mechanisms that lead to tumorigenesis,

are induced and modified differently in HR-HPV than in BPV-1/-2 (Moody & Laimins, 2010).

miRNAs are non-coding small RNAs with essential roles in a great variety of cellular processes and play an important role in HPV oncogenesis (Gómez-Gómez *et al.*, 2013; Banno *et al.*, 2014; Sharma *et al.*, 2014; Wang *et al.*, 2016). Certain miRNAs have been shown to be consistently dysregulated in HPV-induced cancer, such as miR-218 and clusters miR-143/145 and miR-302~367. These miRNAs participate in inhibition of apoptosis (miR-218 and miR-143/145 cluster) and metastasis pathways (miR-218 and miR-302~367 cluster), and during HPV induced cancer they are downregulated by HPV E6 via inhibition of p53 pathways (Wang *et al.*, 2008; Li *et al.*, 2011; McBee *et al.*, 2011; Wald *et al.*, 2011; Ben *et al.*, 2015). Other miRNAs have been reported to be upregulated during HPV tumorigenesis, such as miR-196a and miR-21. These two miRNAs participate in migration, invasion pathways and appear to be dysregulated due to the effect of HR-HPV E7 on PI3K/Akt signalling pathway (Pereira *et al.*, 2010; Gocze *et al.*, 2013; Wang *et al.*, 2016). Nevertheless, several studies have shown discrepancies in the expression of some miRNAs in both *in vitro* and *ex vivo* studies of HPV-induced cancer and this was initially highlighted by Pereira *et al.* (2010) and recently discussed in the review by Sharma *et al.* (2014). Examples of this are hsa-let-7b (Wang *et al.*, 2008; Li *et al.*, 2011), miR-146a and miR-146b-5p (Wang *et al.*, 2008; Lajer *et al.*, 2012), which have shown different levels of expression across studies. Variation in the control samples used in *ex vivo* studies and in clinical samples may be a reason for these disparities (Pereira *et al.*, 2010; Sharma *et al.*, 2014). Additionally, expression of miRNA changes over time after HPV transfection, and therefore differences found in the aforementioned studies, may be due to the dynamic changes occurring during viral infection and neoplastic development (Greco *et al.*, 2011; Li *et al.*, 2011). A summary of miRNAs that have been linked to expression of oncoproteins E5, E6 and E7 in HPV-16/-18 is displayed in Table 5.1.

Table 5.1 List of miRNAs dysregulated in HPV cancer linked to expression of oncoproteins E5, E6 and E7

miRNA	Oncoprotein	Pathway	Level of expression in HPV+	Effect of miRNA dysregulation	Source
miR-106b	E6/E7	E2F1	up	Cell migration	McBee *et al.* (2011); Gómez-Gómez *et al.* (2013)
miR-129-5p	-	Targets E6 via SP1	down	Decreases E6 and increases p53	Zhang *et al.* (2013)
miR-129-5p	-	Targets E7	down	Inhibits E7 expression	Zhang *et al.* (2013)
miR-129-5p	E6/E7	Unknown	down	Dysregulation of the cell cycle	Gómez-Gómez *et al.* (2013)
miR-145	E6	p53	down	Prevents apoptosis	Shi *et al.* (2012)
miR-146a	E5	Unknown	up	Decrease of target PDZD2. Increases dramatically after HPV transfection due to HPV16 E5 protein. Consequence is decrease in p38 and ERK1. miR-146a is involved in cytokine signalling.	Greco *et al.* (2011); Gómez-Gómez *et al.* (2013)
miR-15/16 cluster	E7	c-Myc	up	Metastasis and angiogenesis	Myklebust *et al.* (2011)
miR-196a	E7	PI3K	up	Promotes cell migration and invasion	Pereira *et al.* (2010)
miR-203	E6	unknown	down	Decreases p53 promoting cell proliferation	McKenna *et al.* (2010)
miR-203	E7	MAPK	down	Dysregulation of the cell cycle	Melar-New and Laimins (2010)
miR-205	E6/E7	pRb?	up	Promotes cell proliferation	McKenna *et al.* (2010)
miR-21	E6	Unknown	up	Promotes increase of phosphorylated Akt inhibiting PTEN (a tumour suppressor) and increasing cell proliferation	Ben *et al.* (2015)
miR-218	E6	LAMB3	down	Promotes tumorigenesis	Martinez *et al.* (2008); Ben *et al.* (2015)
miR-23b	E6	p53	down	Increases uPA and promotes cancer development	Yeung *et al.* (2011)
miR-24	E6/E7	Unknown	up	Promotes cell proliferation	McKenna *et al.* (2010)
miR-324-5p	E5	Unknown	down	Increase of target N-cadherin. It is a negative regulator of the oncogenic Hedgehog pathway	Greco *et al.* (2011); Gómez-Gómez *et al.* (2013)
miR-34a	E6	p53	down	Cell cycle disruption that leads to cell transformation	Wang *et al.* (2008); Li *et al.* (2010)
miR-375	-	targets E6	down	Decreases E6 and increases p53	Li *et al.* (2011); Jung *et al.* (2014)
miR-375	-	targets E7	down	Inhibits E7 expression	Li *et al.* (2011); Jung *et al.* (2014)
miR-92	E6	PTEN protein	up	Tumour development	Yu *et al.* (2013)

Using microarray analysis, differences in miRNA expression were found in BPV-1 *in vitro* transformed equine primary fibroblasts (S6-2) compared to equine primary fibroblasts (EqPalF) (Chapter 3). In the current study, differences in miRNA expression in two equine cell lines containing the BPV-1 viral genome and equine primary fibroblasts were explored in more depth using high throughput sequencing (HTS). HTS has become the tool of choice for quantification and annotation of miRNAs due to the ability to discover novel miRNAs, as previous knowledge of the sequences is not necessary, and to the high dynamic range that allows identification of less abundant miRNAs. Using HTS, a catalogue of the miRNA profile in equine cells using equine primary fibroblast was performed (detailed in Chapter 4). In this chapter, the differences in miRNA expression between three cell lines from an equine sarcoid model developed in the University of Glasgow (Yuan *et al.*, 2008a) were studied with the use of HTS to contribute to the understanding of the tumorigenesis mechanisms of BPV-1/-2 oncoproteins in equine sarcoids.

5.2. Materials and methods

5.2.1. RNA extraction, quantitation and assessment of RNA integrity (RIN)

Chapter 2 (General materials and methods) contains details of RNA extraction and RNA quality/quantity assessment. In summary, RNA was prepared using a phenol extraction method (with Trizol®) from equine primary fibroblasts (EqPalF cells), equine primary fibroblasts transformed *in vitro* with BPV-1 genomes (S6-2 cells) and equine fibroblasts obtained from an equine sarcoid tumour (EqSO4b cells) maintained in culture (Yuan *et al.*, 2008a). Features of the cell lines is displayed elsewhere (Table 2.4, Chapter 2). In summary, EqPalF cells do not contain the BPV-1 genomes or express viral proteins whereas S6-2 cells are EqPalF cells *in vitro*-transformed with 2 BPV-1 genomes per cell and EqSO4b are equine fibroblasts obtained from naturally occurring equine sarcoids that contain 8 BPV-1 genomes per cell (Yuan *et al.*, 2008a). Both S6-2 and EqSO4b cells express BPV-1 E5 protein (Sw I variant) (Yuan *et al.*, 2008a). Quantity and integrity of the extracted RNA were assessed and samples with RIN values greater than 9.0 were utilised for sequencing. One RNA sample was obtained

from EqPalF cells and the RNA was divided in three to construct three libraries. For EqSO4 and S6-2 cells, RNA was extracted from three individual cultures and used to prepare three independent libraries for each cell line. The experimental design is summarised in Figure 5.1.

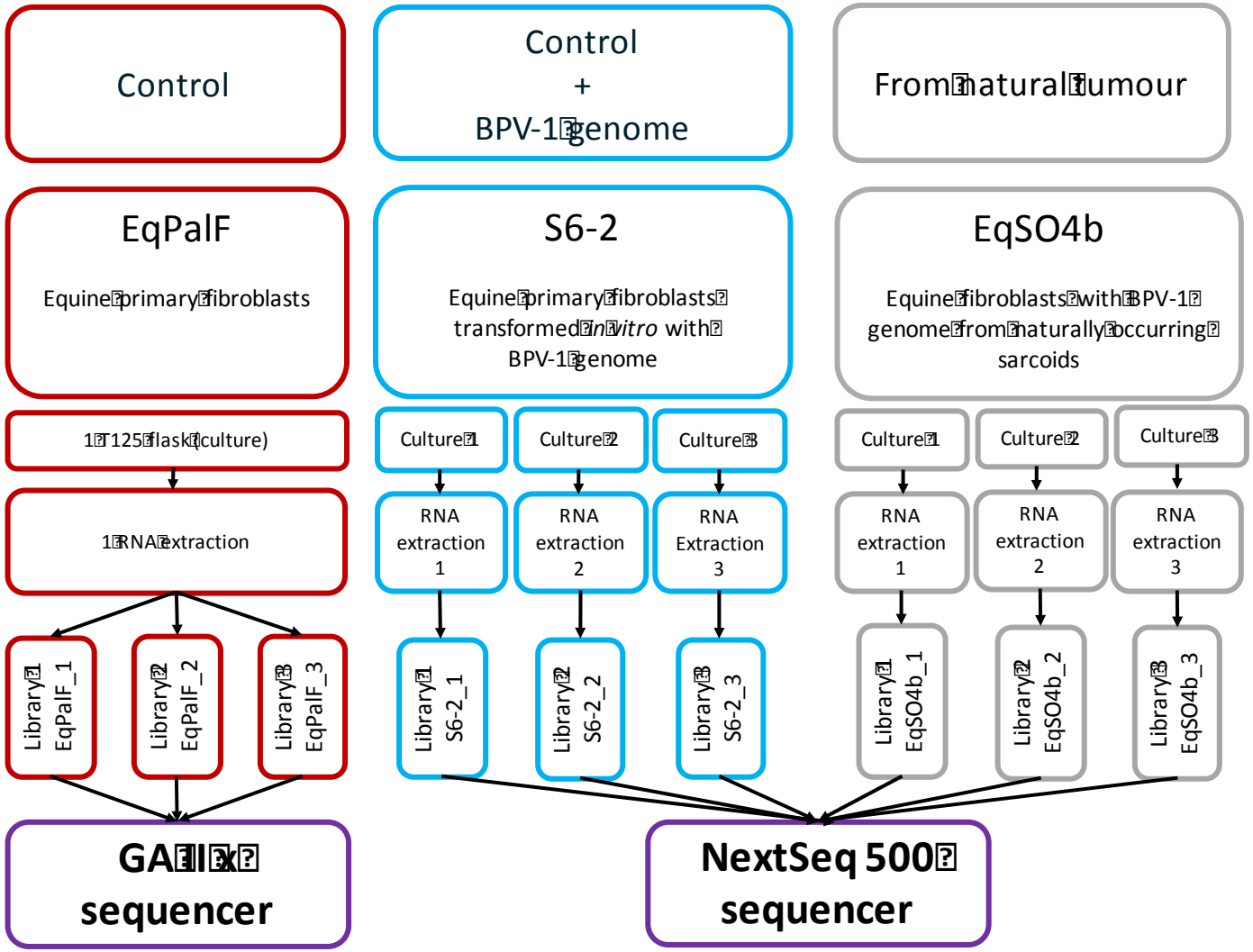


Figure 5.1. Diagram of the HTS experimental design. RNA from each cell line of the equine sarcoid *in vitro* model was obtained, followed by library preparation and sequencing in two platforms (GAIIx and NextSeq500)

5.2.2. Library construction, cluster generation and sequencing

A detailed explanation of library construction, cluster generation and sequencing is given in Chapter 2, section 2.5.1. In summary, libraries were prepared using TruSeq® Small RNA (Life Technologies). Library preparation involves adaptor ligation and size selection. At two different stages of the library construction, the TruSeq® Small RNA kit requires quality control checks to quantify the fraction of potential miRNAs. If the check-points fail, the library preparation is repeated and if enough miRNA fraction is recorded by the DNA high sensitivity chip, the library is highly likely to produce enough fraction to retrieve good quality data. After library construction and cluster generation, massive parallel sequencing was performed in the Genomic Analyzer IIx from Illumina® (GA-IIx) for EqPalF cells and in the NextSeq 500 from Illumina® for EqSO4b and S6-2 cells. The reason for the change of sequencer was due to discontinuity of the use of the GA-IIx machine and reagents. All sequencing was carried out at the University of Glasgow Polyomics facility.

5.2.3. Data analysis

Analysis of the high throughput sequencing data is fully explained in Chapter 2, section 2.5.2. and a summary is displayed in Figure 5.2.

In brief, the files obtained from the GA-IIx and NextSeq500 sequencers were pre-processed (removal of adaptor sequences) and fed into miRDeep 2, software designed specifically for microRNA discovery. During miRDeep2 processing, sequences were mapped against the equine genome (EqCab.2.0, GenBank assembly accession reference GCA_000002305) and equine miRNAs from miRBase 20.0, redundant sequences identified and sequences smaller than 18 nucleotides in length were removed. Following this, RNA folding energy potential was used to identify putative novel miRNAs. The output from miRDeep2, containing the list of sequences of mature miRNA candidates, their genomic location, precursor sequences and number of reads, was harmonised across replicates. To harmonise the data and generate a list of unique miRNA candidates, the following criteria were applied:

1. If two mature sequences were encoded on the same chromosomal location with a maximal deviation of location of 10 nucleotides, they are defined as the same microRNA (tolerance for chromosomal location 10).
2. If two mature sequences were the same length and had the same nucleotide sequence they were considered the same microRNA (i.e. tolerance for sequence composition 0).

The genomic location condition was applied first and nucleotide sequence followed. This was to prevent discarding the historically named 'star miRNA' (miRNA* or passenger strand), as miRDeep2 gives priority to the guide strand Each potential microRNA candidates was assigned an internal ID in the form of U_*number_location in the precursor* (3p or 5p). miRNA candidates were then filtered during post-processing to increase robustness of the data, so that only the miRNAs present in at least two of the three replicate libraries of each cell line were included in the analysis.

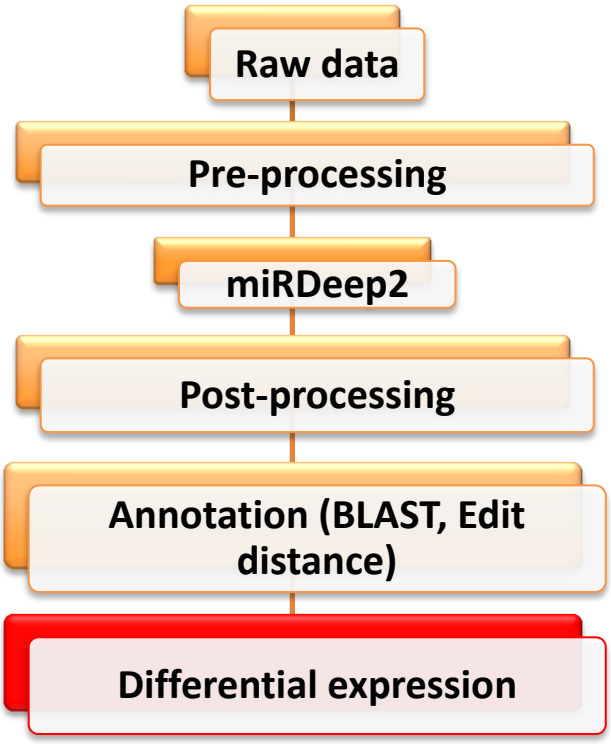


Figure 5.2. Summary of the steps of data analysis of HTS data

To verify if the miRNAs were novel or had been described before, the list of miRNA candidates were analysed using two methods, BLAST + and Edit distance using the list of miRNAs from miRBase 20.0 and the list of equine miRNAs available in the literature. Full details of both methods are explained in Chapter 2, section 2.5.2.2.

To study the differences of miRNA profile between cell lines, the data from the three cell lines (9 replicates in total) was analysed using DESeq R package (Anders & Huber, 2010) from Bioconductor (http://bioconductor.org/packages/release/bioc/html/DESeq.html). With this package, firstly reads from each cell line were normalised to the level of the sample with the highest read number and the differences in fold change between the cell lines were calculated in a pair-wise manner (EqPalF *vs.* EqSO4b, EqPalF *vs.* S6-2 and EqSO4b *vs.* S6-2 cells). The p-value for differential expression of each putative miRNA was calculated; following this, an adjusted p-values was also obtained using Benjamini-Hochberg procedure to control for multiple tests (Anders & Huber, 2010). An adjusted p-value of less than 0.05 was taken to be statistically significant. Data was subsequently explored using heat maps, MvA plots (also known as Bland-Altman) and principal component analysis (PCA). Figure showing the data for differential expression analysis were constructed in Microsoft Excel, ClustVis and R. Functional analysis was performed using Ingenuity Pathway Analysis (Qiagen) where identified miRNAs could be mapped to their human or mice equivalent. Biomarker analysis was performed using LefSe, implemented using Galaxy.

5.2.4. Validation using real-time quantitative Reverse-Transcription Polymerase Chain Reaction (qRT-PCR) with Taqman technology

Real-time Quantitative Reverse-Transcription Polymerase Chain Reaction (qRT-PCR) was used to quantify differential expression for a subset of miRNAs identified as differentially expressed by the HTS approach. Six miRNAs were selected as they fulfilled all the following criteria:

1. Differentially expressed between EqPalF *vs.* EqSO4b, EqPalF *vs.* S6-2 or EqSO4b *vs.* S6-2 cells ($p<0.01$) and present in the top 60 miRNAs ranked by absolute fold change in each pair-wise comparison. When a miRNA was present in more that one comparison, only one entry for that miRNA was kept.
2. Showed a perfect match to equine miRNAs included in miRBase using BLAST (see Chapter 2 for more details), setting the values of identity at 100% and coverage at 100%.
3. Commercial availability of Taqman® probes for the miRNA sequence (Life technologies).
4. Biological significance (i.e. having being previously reported to be dysregulated in cancer and/or HPV).

5.3. Results

5.3.1. RNA quality, integrity and library quality control

Total RNA extracted from EqPalF (equine primary fibroblasts), S6-2 (equine primary fibroblasts *in vitro* transformed with BPV-1 genomes) and EqSO4b cells (equine fibroblasts from naturally occurring sarcoids) maintained in culture showed good integrity as observed by electrophoresis on 1% agarose gel. The A260/A280 ratio ranged from 2.02 to 2.07 and the RIN value ranged from 9.90 to 10.0, obtained using the Bioanalyzer 2100, indicating suitable RNA for sequencing (see Figure 4.3 for details on EqPalF and see Figure 5.3 for details on EqSO4b and S6-2 RNA). During library preparation, EqSO4b and S6-2 RNA samples shown a discrete peak on the DNA chip as expected for the miRNA content, whereas one of the libraries from one of the replicates from EqPalF had to be repeated due to suspected contamination (see Chapter 4, section 4.3.1.).

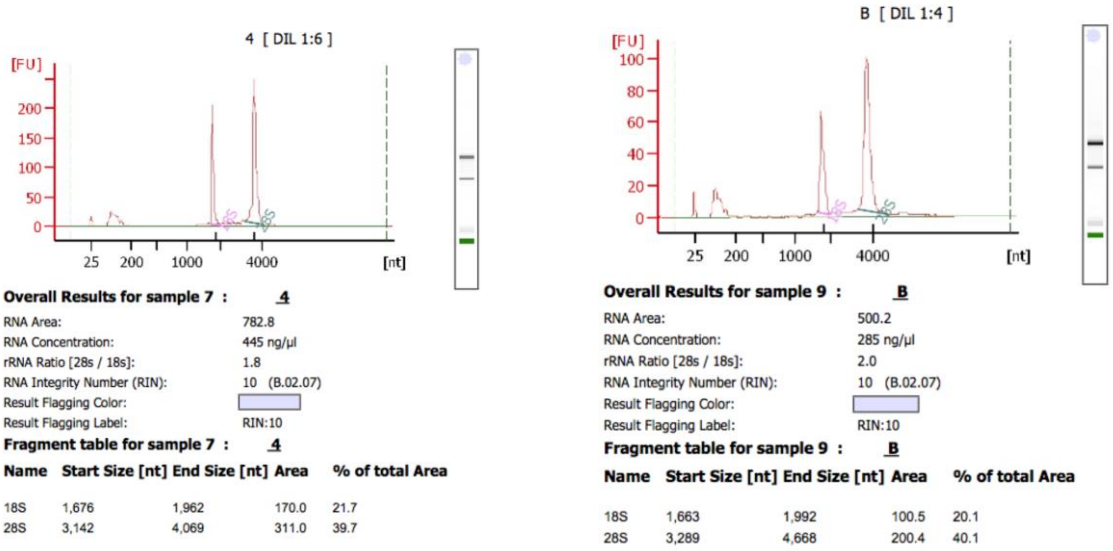


Figure 5.3. Electrophenogram obtained with the Bioanalyzer 2100 of RNA samples from EqSO4b (left) and S6-2 (right). No signs of RNA degradation were noticed and RIN values were 10 for both RNA samples

5.3.2. Characterisation of RNA read data - read count, number of mature miRNA candidates and correlation across replicate libraries

The number of reads obtained from EqPalF cells from the GAIIx sequencer ranged from 7.81 x 10^6 - 10.70 x 10^6 whereas reads obtained from EqSO4b and S6-2 cells from NextSeq500 ranged from 22.4 x 10^6 – 29.5 x 10^6. The differences in read counts between EqPalF and the other two cell lines can be explained by differences in the sequencing platforms. Sequencing reads after adaptor-trimming were fed into miRDeep2. This mapped the sequencing reads to the databases (equine genome and equine miRNAs included in miRBase 20.0), removed redundancy and performed miRNA predictions based on a core algorithm and folding energy of the miRNA precursor candidates. Sequences that were not mapped to the equine genome, mapped to other types of RNA such as tRNA, mapped to five or more loci in the genome, and sequences that did not meet the folding energy criteria set for miRNA in the Randfold script were removed. The number of reads retained after miRDeep2 and harmonization ranged from 3.82 x 10^6 – 6.72 x 10^6 in EqPalF, 14.2 x 10^6 – 18.0 x 10^6 in EqSO4b and 12.0 x 10^6 - 15.1 x 10^6 in S6-2 cells. Combining the output of the three cell

lines reads were normalised to 20 x 10^6 reads in each cell line. A Perl script was used to harmonise the data between cell lines and 1336 potential miRNA candidates were found. The list of 1336 miRNA candidates is included in the file 'master file DESeqdata.xlsx' in the DVD enclosed. Harmonisation of the data was performed with the following assumptions: two sequences were considered the same mature miRNA if they were in the same genomic location (with a maximum deviation of ten nucleotides) and if the nucleotide composition was exactly the same (tolerance 0 for nucleotide sequence). When these criteria were followed and there were differences in length either at the 3' or 5' end, the longest sequence was used for further analysis. To add robustness to the results, the list of mature miRNA candidates was filtered by occurrence. All miRNAs that were sequenced in at least two of three the libraries of each of the three cell lines, independently of the read count, were included for further analysis. The list was therefore shortened from 1336 to 752 mature miRNA candidates. In total, 593, 558 and 585 mature miRNAs were found in EqPalF, EqSO4b and S6-2 cells respectively. Of those, 433 miRNAs were commonly expressed in the three cell lines, 84 were exclusive to EqPalF, 65 and 52 were exclusive to EqSO4b and S6-2 respectively (see Figure 5.4.).

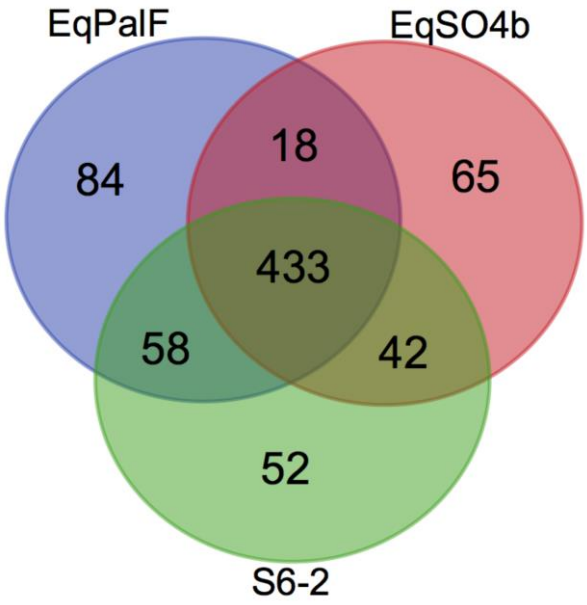


Figure 5.4. Venn-diagram indicating the number of miRNAs present in each cell line. EqPalF expressed 593 miRNAs while EqSO4b and S6-2 expressed 558 and 585 respectively. A substantial number of miRNAs (433) was present in all three cell lines. Presence of a miRNA in each cell line was defined as 'at least present in two of the three replicates of that cell line'

Pearson's correlation coefficient (r) was 0.96-1.00 for EqPalF replicates, 0.98-0.99 for EqSO4b replicates and 0.99 for S6-2 replicates, indicating that the replicates were very similar in miRNA expression pattern and miRNA repertoire within each cell line, as the values are closer to 1 rather than closer to 0. As r coefficient is higher than 0, the direction of the linear relationship is positive in all cell lines. Graphical representation of the correlation of miRNA expression between replicates from S6-2 and EqSO4b is displayed in Figure 5.5. Correlation of replicates for EqPalF was displayed elsewhere (Chapter 4, section 4.4.2., Figure 4.4.). After normalisation, data was also assessed using MA plots (Blant-Altmant plots), whereby differences in expression of miRNAs in each cell line pair were plotted against average expression. Figure 5.6. shows the distribution of the data of cell lines in pairs. The most differentially expressed miRNAs are located to the higher side of the x axis (average expression) and away from "0" on the y axis (Log_2FC of expression).

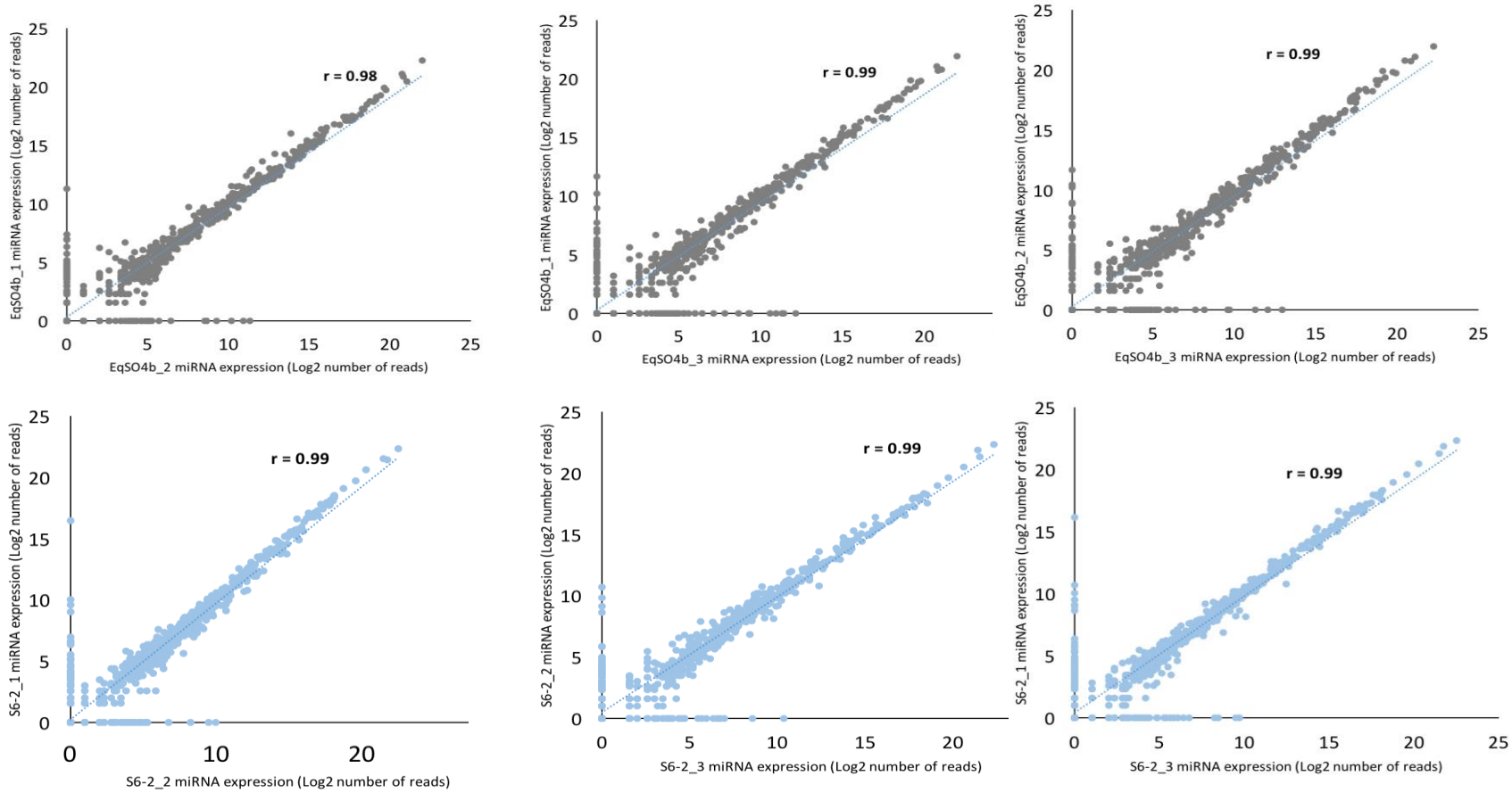


Figure 5.5. Correlation of miRNA expression in EqSO4b (above) and S6-2 cells (below). The number of reads of each miRNA, expressed in Log_2 values from each of the three replicates, was compared pair-wise. Each dot represents one miRNA with the Log_2 read number in x axis from one replicate and in y axis from a different replicate. Pearson's coefficient for each pair is displayed as 'r' and shows a positive and strong linear relationship in the three pair comparisons (as values are close to +1.0)

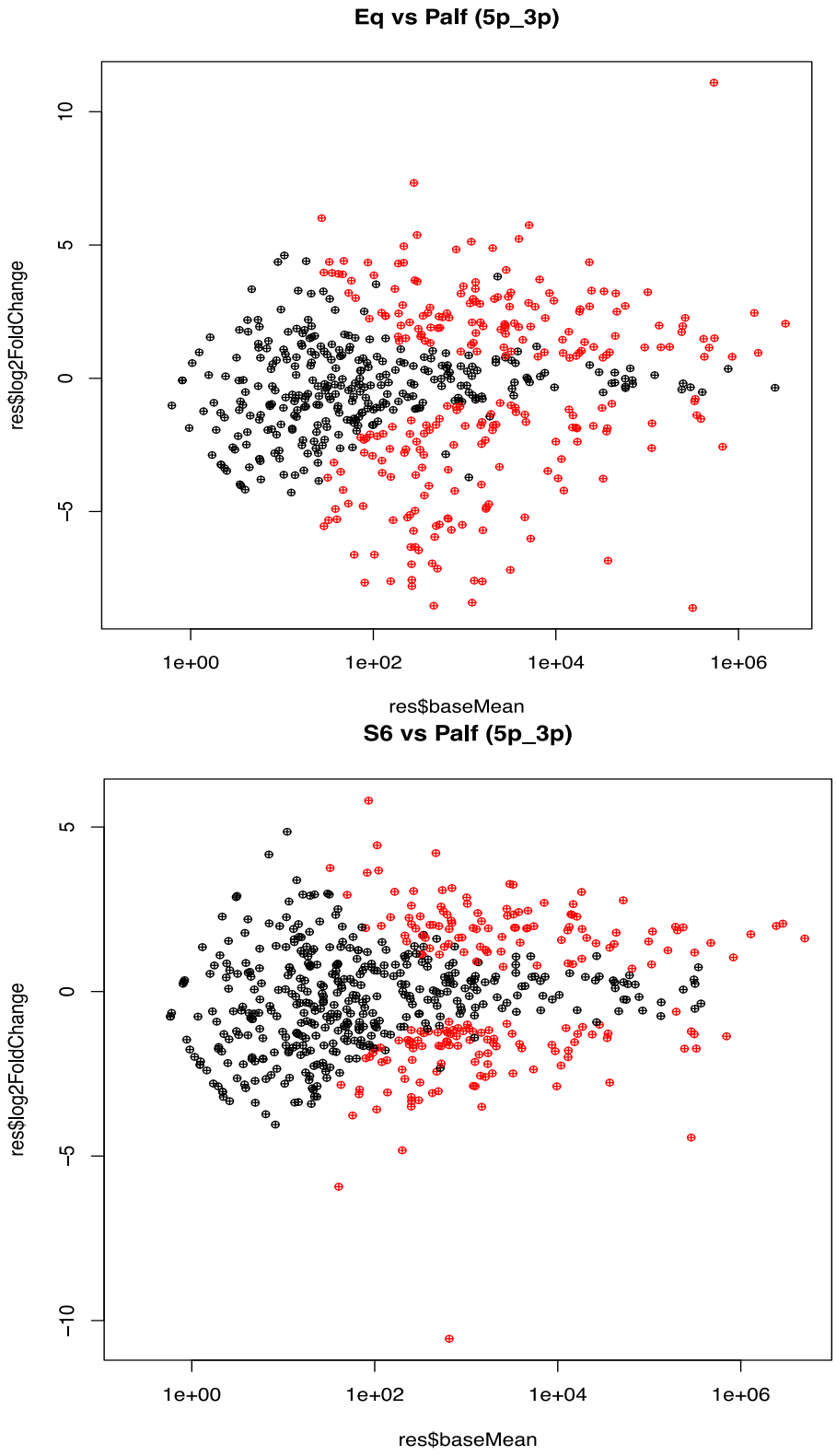


Figure 5.6. MA plots of miRNA expression in EqPalF vs. EqSO4b (above) and EqPalF vs. S6-2 (below). miRNAs that were differentially expressed with an adjusted p value <0.05 are represented in red

The difference in the miRNA expression levels was also small within each cell line as observed in the PCA plots. This was particularly noticeable in EqPalF replicates 1, 2 and 3 where the replicates sit very close to each other in the PCA plot (see Figure 5.7). EqSO4b and S62 dissimilarities were also small but not as small as EqPalF. This may be explained by the fact that the three EqPalF libraries were prepared from the same RNA preparation. Nevertheless, the majority of the differences were found between each cell line: almost 84% of the variation between EqPalF and EqSO4b, and almost 80% of the variation between EqPalF and S6-2 is due to the inherent biological differences between cell lines (see Figure 5.8).

Differences between the three cell lines were explored using a three-way PCA plot with the list of miRNAs expressed in each cell line and levels of expression. Differences between the cells containing BPV-1 genomes were smaller when compared with each other and were bigger when compared to EqPalF cells (see Figure 5.8). Almost 50% of the variation in miRNA expression between EqPalF and cells containing BPV-1 genomes (S6-2 and EqSO4b) can be attributed to the biological differences in the cell lines due to the presence of BPV-1 viral genomes.

Out of the 752 miRNA candidates found, 280 had an equine match in miRBase 20.0 (eca-miR-xxx, where "x" are the number or letters that have been used to annotate each miRNA), 187 had a match in miRBase 20.0 from other species (of which, 80 were a perfect match to human miRNAs or hsa-miR-xxx) and 285 had no match in the database and therefore are considered novel miRNAs to the horse and to other species. Novel miRNAs, which annotation is referred to as U_number_-5p/3p in this chapter, were excluded from any functional analysis as for the analysis either miRBase annotation name (eca-miR-xxx) or miRBase number (MIMAxxxxxxx) is required. For details on those miRNAs (i.e. mature sequence, levels of expression, precursor sequence, genomic location, DESeq data) please see file 'master file DESeqdata.xlsx' in the accompanying DVD.

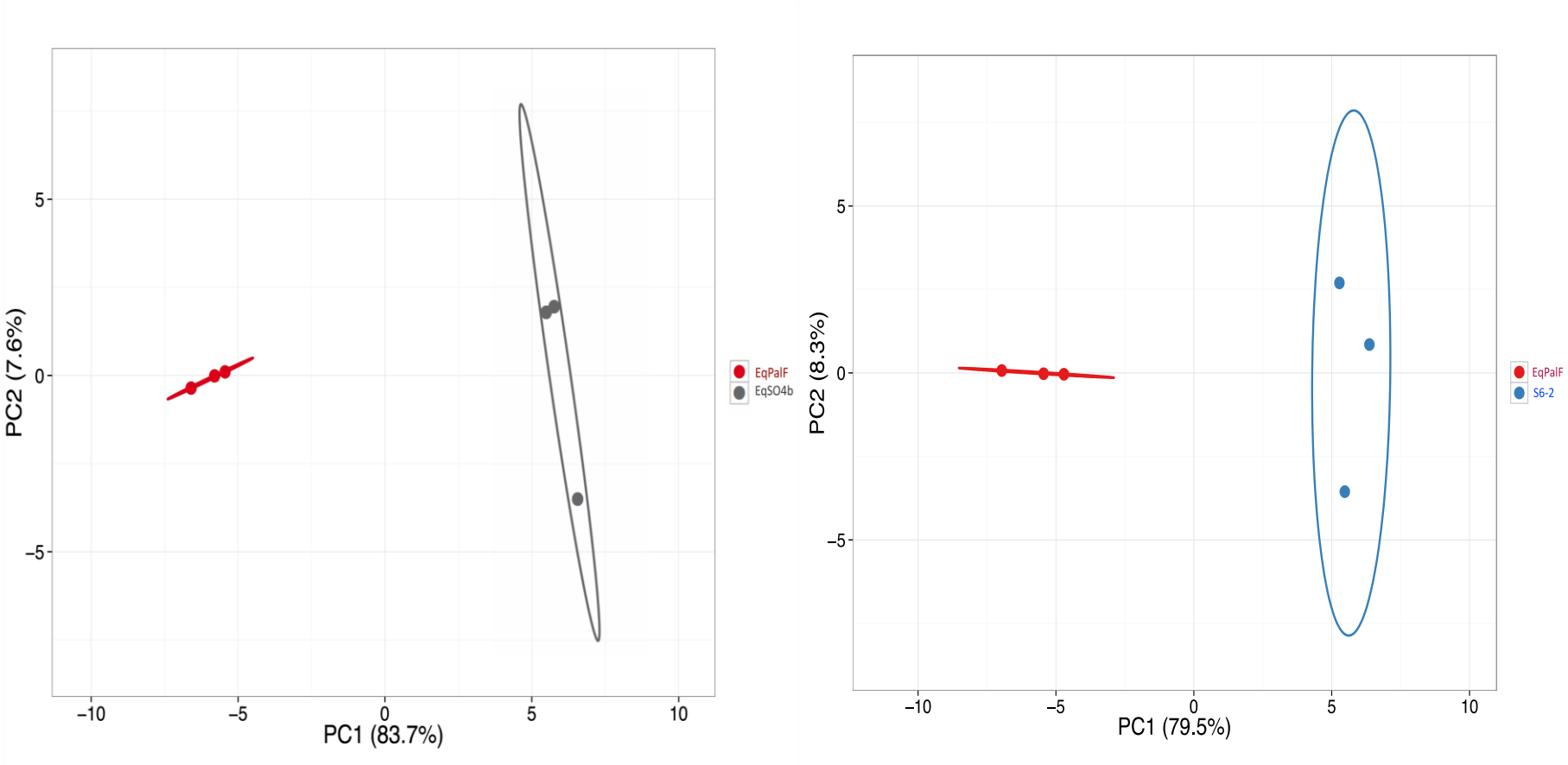


Figure 5.7. Principal Component Analysis (PCA) plot of list of miRNAs, their expression in number of reads and presence in each cell line (pair comparisons). Unit variance scaling is applied to rows (3 replicates per cell line). Singular value decomposition (SVD) with imputation was used to calculate principal components. Axis x and y show principal components 1 and 2 that explain 79.5% and 8.3% of the total variance respectively in S6-2 vs. EqPalF (right) and explain 83.7% and 7.6% of the total variance respectively in EqSO4b vs. EqPalF (left). Prediction ellipses are such that a new observation from the same cell line would fall inside the ellipse (based on 6 data points, probability 0.971)

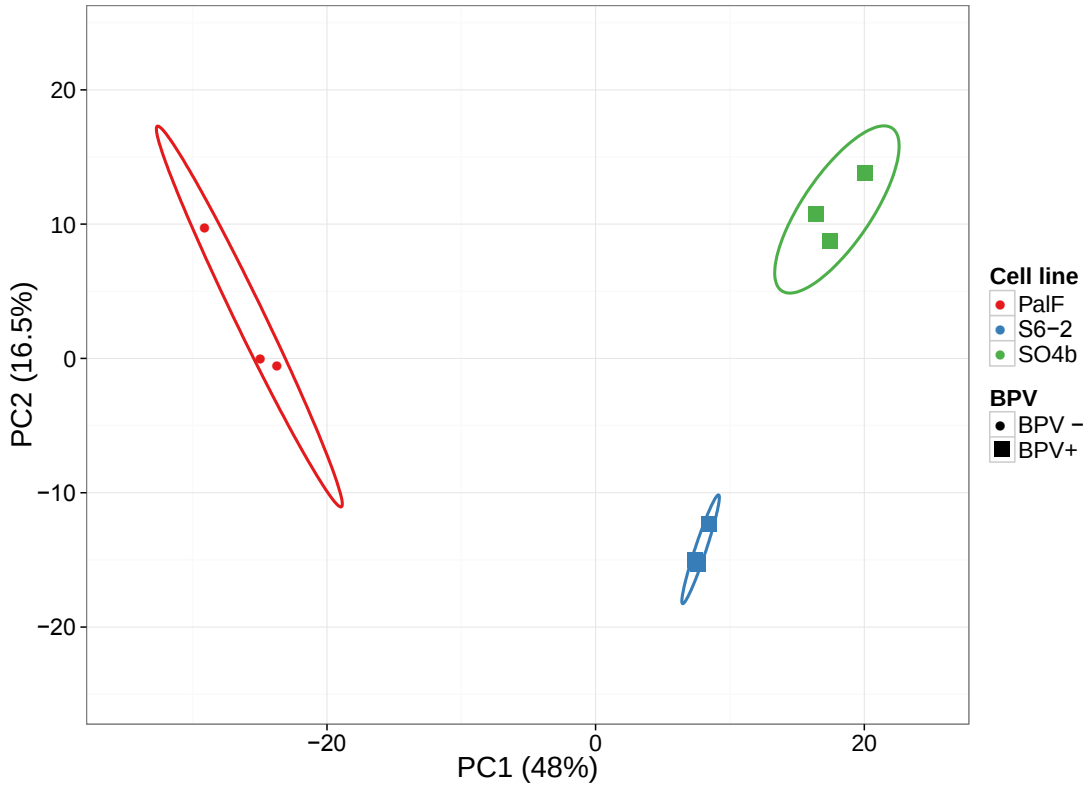


Figure 5.8. Principal Component Analysis (PCA) plot of list of miRNAs, their expression indicated by number of reads and presence in the three cell lines (three-way comparison). Singular value decomposition (SVD) with imputation was used to calculate principal components. Axis x and y show principal components 1 and 2 that explain 48% and 16.5% of the total variance respectively (based on 9 data points, probability 0.92).The plot demonstrates that the biggest differences between the cell lines are across PC1, showing that S6-2 cells are closer to EqSO4b cells than to EqPalF cells

5.3.3. Differential expression of miRNAs between equine primary fibroblasts *in vitro* transformed with BPV-1 genomes (S6-2 cells) and equine primary fibroblasts (EqPalF)

DESeq results

In total, 251 mature miRNA candidates were found significantly dysregulated in S6-2 (fibroblasts transformed in vitro with BPV-1 genome) compared to EqPalF (primary fibroblast) cells (adjusted $p<0.05$). Of these differentially expressed miRNAs, 19% (49) have not been previously reported neither in the horse nor any other species and are therefore considered novel miRNAs. Novel miRNAs were not included in the functional analysis because of the reasons mentioned in section 5.3.2. Just over half of the differentially expressed miRNAs (52%, 131 miRNAs) were found to be upregulated in cells containing the BPV-1 genomes. Differences in expression of miRNA candidates between S6-2 and EqPalF cells ranged from 0.61-10.82 (Log_2FC). The top 10 most up - and down-regulated miRNA candidates in S6-2 compared to EqPalF are displayed in Table 5.2. and their differences are presented in Figure 5.9. Interestingly, 14 of those had no match in miRBase 20.0 (novel miRNAs). In addition, in the top 50 most dysregulated miRNA candidates, 31 miRNAs in total did not have a match in the miRBase. While this limited the use of the data for automated functional analysis, it is an interesting finding. The two most dysregulated miRNAs were U147_5p (novel miRNA) and hsa-miR-1249-3p (U140-3p) with Log_2FC of 10.82 and -10.55 respectively (see Table 5.2). The top 50 most dysregulated miRNA candidates (top 25 upregulated in S6-2 and top 25 downregulated compared to EqPalF) are presented in Figure 5.10). Hierarchical clustering positions the three replicates from EqPalF together and the three replicates from S6-2 together, and highlights a difference between the two cell lines as shown by the separation of the branches of the clustering tree.

Table 5.2. List of the 10 most upregulated and 10 most downregulated miRNAs in S6-2 when compared to EqPalF ranked by log_2FC. When no match for a miRNA was found in the database (miRBase) the internal identifier generated after harmonising the data, was used instead (U**number_3p/5p). Log_2FC represents the difference in expression between the two cell lines calculated as the average expression in each cell line after normalising read counts to 20 x 10^6 reads. Fourteen of the 20 most dysregulated miRNA candidates have no annotation in the miRBase 20.0 and therefore are novel miRNA candidates

ID	Mature sequence	Length	Annotation	p-adj	Log 2 fold change
U147_5p	AUCCUGCCGACUGCGCCA	18	U147_5p	3.10E-92	10.82
U174_3p	CUAGAAUGAAGCUCCUAGAGG	21	U174_3p	2.56E-63	9.95
U142_5p	UAGGGGUAUGAUUCUCGCU	19	U142_5p	7.96E-06	9.89
U201_3p	UUGCCUGGGCCUCUGGAACCAU	22	U201_3p	7.99E-58	9.76
U183_3p	AAGUUUCUCUGAACGUGUAGAGC	23	U183_3p	1.27E-05	9.75
U207_3p	UCAAGGAGAUCACAGUCUAGU	21	U207_3p	6.43E-41	9.11
U230_5p	UCAGUAACAAAGAUUCAUCCUUG	23	eca-miR-802	8.96E-40	9.05
U238_5p	UGAGAUGAAGCACUGUCAC	19	U238_5p	5.94E-05	8.43
U301_5p	AGCCAGCUUGAGCCCACCUACAU	23	U301_5p	3.63E-08	6.51
U316_3p	ACAAAAGUCUGCACAUUGGA	20	U316_3p	8.29E-04	6.49
U279_3p	CUGACCCCCUGCCCCCCCCAUC	22	U279_3p	1.66E-04	-5.86
U310_3p	UGGACGGAGAACUGAUAAGGGU	22	eca-miR-184	1.15E-04	-5.91
U290_5p	UGGCUCUCCUCUGCCUGCUGUGC	23	U290_5p	1.27E-04	-5.93
U286_3p	GACCCCCAGCCCCAGCCCUGGA	22	U286_3p	3.46E-05	-6.07
U277_5p	UGAGAACUGAAUUCCAUGGGUU	22	eca-miR-146a	9.60E-07	-6.48
U263_3p	UCUGCCUCUUGUUUCCUUCCAGC	23	U263_3p	3.50E-03	-6.49
U275_3p	AUUUCCACCACGUUCCCG	18	U275_3p	3.27E-09	-6.95
U154_5p	GAGGGCCCCCCCCAAUCCUGU	21	ssc-miR-296-5p	2.87E-20	-8.88
U154_3p	AGGGUUGGGUGGAGGCUUUCCU	22	eca-miR-296	8.16E-41	-9.39
U140_3p	ACGCCCUUCCCCCCCUUCUUCA	22	eca-miR-1249	3.50E-22	-10.55

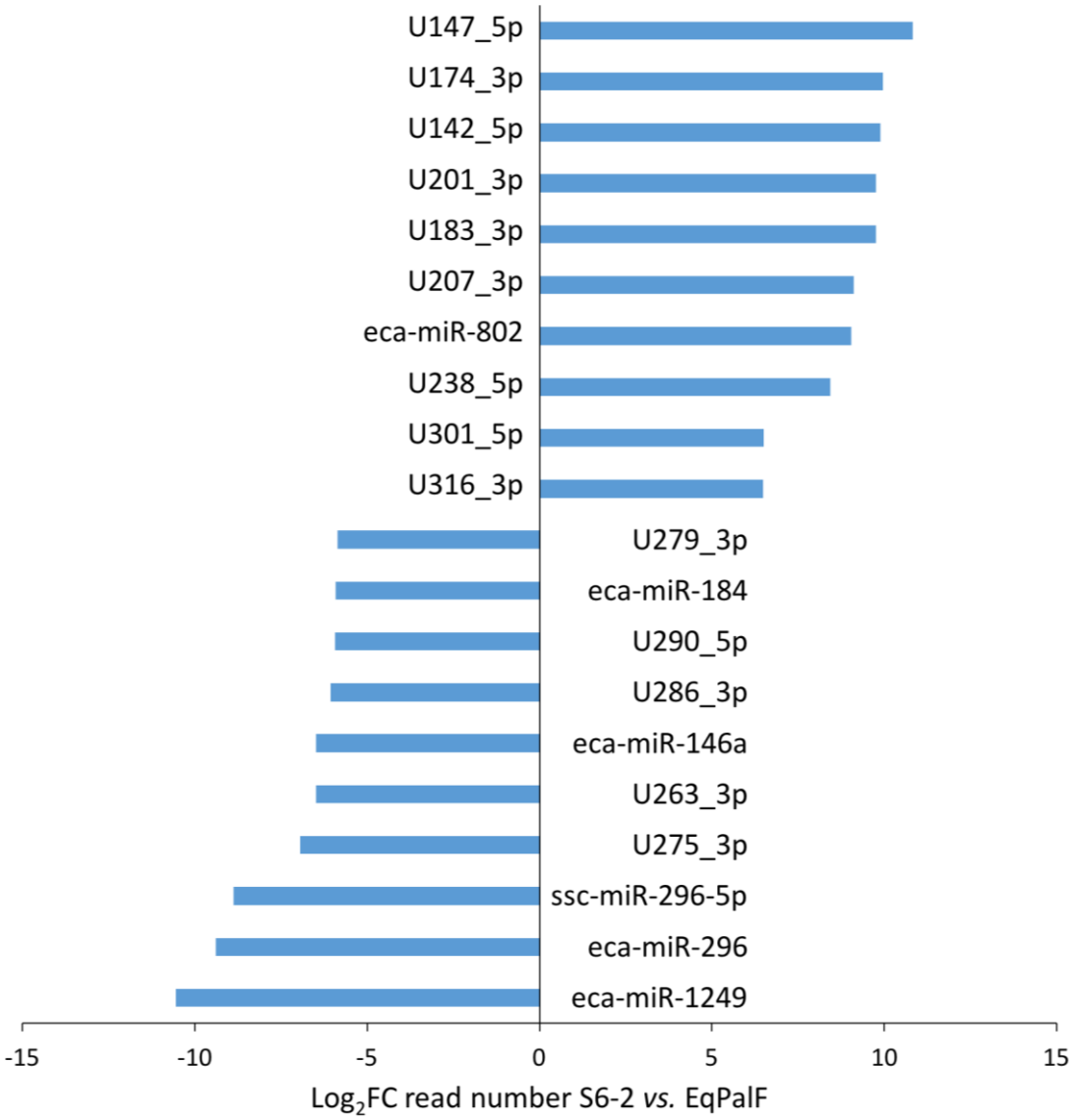


Figure 5.9. Top most differentially expressed miRNAs in S6-2 vs. EqPalF (p adjusted <0.05) ranked by Log_2FC (ten most upregulated and ten most downregulated). When no match was found in miRBase, the miRNA maintained the internal annotation (U_number_3p/5p). The x-axis denotes the Log_2FC between reads from S6-2 and EqPalF: overexpressed miRNAs in S6-2 have values >0 and underexpressed miRNAs have values <0

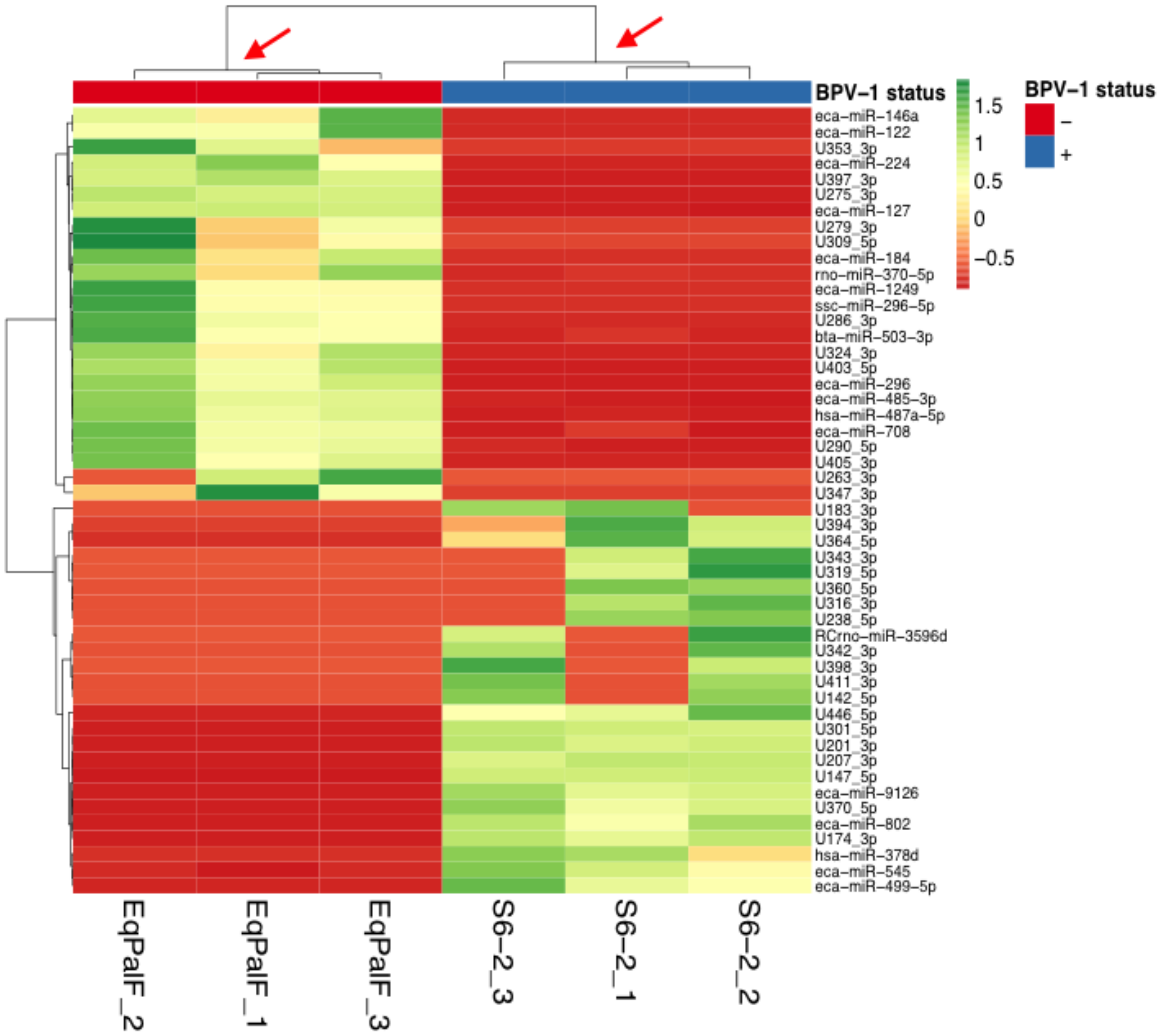


Figure 5.10. Heat map and hierarchical clustering of the top 50 most differentially expressed miRNAs in S6-2 vs. EqPalF (Adjusted p<0.05) ranked by Log_2FC (25 most upregulated and 25 most downregulated). When no match for a miRNA was found in the database (miRBase) the internal identifier generated after harmonising the data, was used instead (U"number"_"3p/5p"). Hierarchical clustering was performed using correlation as clustering distance with average as the linkage criteria; the tree was ordered by the tightest cluster first. Each column represents one replicate (EqPalF_1,_2 and _3 for EqPalF and S6-2_1,_2 and _3 for S6-2). Note that through hierarchical clustering the replicates from each cell line are closer to each other than to the counterpart cell line, shown by the short branch lengths within both clusters (red arrows)

Comparison of DESeq data with miRNA microarray

Of the 251 miRNAs found dysregulated in S6-2 cells when compared to EqPalF, 42% (105 miRNAs) had been previously interrogated using microarray (for details see Chapter 3). Of those, just over half had shown signal intensity above the background level during microarray (55%, 57 miRNAs) and the rest were found below the background level (45%, 48 miRNAs). Out of the 57 with signal above background level, 37% (21 miRNAs) showed the same pattern of expression between HTS and microarray (either upregulated or downregulated), 18% (10 miRNAs) were not differentially expressed by microarray but were on the HTS dataset and the rest (26 miRNAs, 46%) showed disagreement in the direction of the dysregulation (up or down). miRNAs found differentially expressed in both HTS and microarray with agreement in the direction of dysregulation, are displayed in Table 5.3. Three of those miRNAs (U72-5p, U4-5p and U44-5p) were found to be encoded in regions of the equine genome previously reported to be linked to equine sarcoids (Jandova *et al.*, 2012; Bugno-Poniewierska *et al.*, 2016).

Table 5.3. List of 21 miRNAs that show agreement in the comparison of S6-2 cells vs. EqPalF by HTS and microarray (p<0.05).
Differences are displayed as Log_2FC between S6-2 and EqPalF. In bold the three miRNAs encoded in regions of the equine genome linked to higher susceptibility to equine sarcoids have been highlighted

HTS					Microarray	
ID	Mature sequence	ECA	Annotation	Log_2FC	Log_2FC	Result (HTS and microarray)
U9_3p	UCGGAUCCGUCUGAGCUUGGCU	24	eca-miR-127	-4.44	-2.13	downregulated in BPV-1 +
U108_3p	GUCAUACACGGCUCUCCUCUCU	24	eca-miR-485-3p	-3.50	-0.53	downregulated in BPV-1 +
U105_5p	UGGUUUACCGUCCCACAUACAU	24	hsa-miR-299-5p	-2.88	-0.35	downregulated in BPV-1 +
U34_3p	GAAUGUUGCUCGGUGAACCCCU	24	eca-miR-409-3p	-2.77	-1.20	downregulated in BPV-1 +
U96_5p	UCUUGGAGUAGGUCAUUGGGUGG	24	eca-miR-432	-2.49	-0.73	downregulated in BPV-1 +
U58_5p	UAGCAGCACAUCAUGGUUUACA	5	eca-miR-15b	-1.99	-1.00	downregulated in BPV-1 +
U72_5p	**UUCCCUUUGUCAUCCUAUGCCU**	**23**	**eca-miR-204b**	**-1.81**	**-0.16**	**downregulated in BPV-1 +**
U64_3p	UAACAGUCUACAGCCAUGGUCG	11	eca-miR-132	-1.79	-1.48	downregulated in BPV-1 +
U10_3p	UAUUGCACUUGUCCCGGCCUGU	17	eca-miR-92a	-1.73	-0.40	downregulated in BPV-1 +
U104_5p	UUAAUGCUAAUCGUGAUAGGGGU	26	eca-miR-155	-1.67	-0.17	downregulated in BPV-1 +
U148_5p	GAAGUUGUUCGUGGUGGAUUCG	24	eca-miR-382	-1.51	-0.53	downregulated in BPV-1 +
U30_5p	UGAGGGGCAGAGAGCGAGACUUU	11	eca-miR-423-5p	-1.41	-0.14	downregulated in BPV-1 +
U4_5p	**AACAUUCAACGCUGUCGGUGAGU**	**25**	**eca-miR-181a**	**-1.36**	**-1.27**	**downregulated in BPV-1 +**
U50_3p	ACAGCAGGCACAGACAGGCAGU	5	eca-miR-214	-1.30	-0.77	downregulated in BPV-1 +
U11_5p	CAACGGAAUCCCAAAAGCAGCUG	16	eca-miR-191a	-1.29	-0.46	downregulated in BPV-1 +
U44_5p	**AACAUUCAUUGCUGUCGGUGGGU**	**25**	**eca-miR-181b**	**-1.27**	**-0.87**	**downregulated in BPV-1 +**
U73_5p	UGGGUCUUUGCGGGCGAGAUGA	11	eca-miR-193a-5p	-1.17	-0.12	downregulated in BPV-1 +
U17_3p	AGCUACAUUGUCUGCUGGGUUUC	X	eca-miR-221	-0.61	-0.32	downregulated in BPV-1 +
U76_5p	AACCCGUAGAUCCGAUCUUGUG	26	eca-miR-99a	1.86	0.35	upregulated in BPV-1 +
U158_3p	UGUGACAGAUUGAUAACUGAAA	X	eca-miR-542-3p	2.03	0.12	upregulated in BPV-1 +
U100_3p	UAGCACCAUUUGAAAUCAGUGUU	4	eca-miR-29b	2.67	0.11	upregulated in BPV-1 +

5.3.4. Differential expression of miRNAs between equine tumoural fibroblasts (EqSO4b) and equine primary fibroblasts (EqPalF)

In total, 329 mature miRNA candidates were found differentially expressed between EqSO4b and EqPalF cells (p adj<0.05). Differences between levels of expressions ranged from -16.5 to 13.34 (measured as Log_2 fold change (Log_2FC) of expression in EqSO4b compared to EqPalF). Of the 329 miRNAs, 82% (271 miRNAs) had an annotation allocated in miRBase 20.0 and 18% (58 miRNAs) had not been reported in the horse or any other species (novel miRNAs). Novel miRNAs were excluded from the functional analysis because of the reasons mentioned in section 5.3.2. Out of the 329 differentially expressed miRNAs, 51% (168 miRNAs) were upregulated in EqSO4b and 49% (161 miRNAs) were found downregulated when compared to EqPalF cells, very similar to the comparison between S6-2 and EqPalF cells presented above. The 20 most dysregulated miRNAs (top 10 upregulated and top 10 downregulated), ranked by Log_2FC, are displayed in Table 5.4. Their differences in Log_2 FC are represented in Figure 5.11. The top most dysregulated miRNAs were eca-miR-23a (U29_3p) and eca-miR-615-3p (U88_3p) with Log_2FC -16.50 and 13.34 respectively (see Table 5.4).

Differences in expression between the top 50 most dysregulated miRNAs (top 25 upregulated and top 25 downregulated in EqSO4b compared to EqPalF), ranked by Log_2FC is displayed in Figure 5.12. It is worth noting that the hierarchical clustering of this heat map puts the replicates of each cell line close to each other as the branches of the clustering tree are tight, whereas there is a marked difference between the two cell lines as seen by the separation of the main branches of the clustering tree. It is also apparent that for less abundant miRNAs, the differences in expression between replicates are visually more obvious in the heat map. This occurs with miRNAs such as U307_5p, U202_3p, U301_5p and efu-miR-219 to name a few. For these miRNAs, the level of expression in all three replicates of EqSO4b is low or cannot be detected in one of the replicates, and therefore the fold change of the average expression of each cell line is indeed significant although there was variation across replicates.

Table 5.4. List of the 10 most upregulated and 10 most downregulated miRNAs in EqSO4b when compared to EqPalF ranked by Log_2FC. When no match for a miRNA was found in the database (miRBAse 20.0) the internal identifier generated was used instead. Log_2FC represents the difference in expression between the two cell lines calculated as the average expression in each cell line after normalising read counts to 20 x 10^6 reads

ID	Mature sequence	Length	Annotation	p adjusted value	Log 2 fold change
U88_3p	UCCGAGCCUGGGUCUCCCUCU	21	eca-miR-615-3p	4.26E-44	13.34
U97_5p	UAGGUAGUUUCCUGUUGUUGGG	22	eca-miR-196b	1.20E-126	13.19
U120_5p	UAGGUAGUUUCAUGUUGUUGGG	22	eca-miR-196a	1.10E-107	12.37
U7_5p	UACCCUGUAGAUCCGAAUUUGU	22	eca-miR-10a	7.32E-145	11.09
U174_3p	CUAGAAUGAAGCUCCUAGAGG	21	U174_3p	3.78E-69	10.82
U178_3p	CACUAGAAUGUGAGCUAC	18	U178_3p	8.62E-06	10.00
U7_3p	CAAAUUCGUAUCUAGGGGAAU	21	oan-miR-10a-3p	5.00E-46	9.82
U201_3p	UUGCCUGGGCCUCUGGAACCAU	22	U201_3p	2.49E-45	9.79
U207_3p	UCAAGGAGAUCACAGUCUAGU	21	U207_3p	1.20E-41	9.61
U197_5p	UAUAUAUAUAUAUGUACGUAUG	22	tch-miR-1277-5p	2.18E-16	9.43
U233_3p	CACAAUACACGGUCGACCUCU	21	U233_3p	4.26E-17	-8.14
U130_3p	CACAUUACACGGUCGACCUCU	21	eca-miR-323-3p	6.35E-60	-8.42
U223_3p	CAACCUGGAGGACUCCAUGCUGU	23	eca-miR-490-3p	2.36E-20	-8.44
U108_5p	AGAGGCUGGCCGUGAUGAAUUC	22	eca-miR-485-5p	9.38E-37	-8.53
U9_3p	UCGGAUCCGUCUGAGCUUGGCU	22	eca-miR-127	4.43E-106	-8.62
U214_3p	AACACACCUAGUUAACCUCUUU	22	eca-miR-329a	6.31E-24	-8.72
U160_5p	UGUGACUGGUUGACCAGAGGGG	22	eca-miR-134	2.09E-05	-9.95
U175_3p	AUCAUGAUGGGCUCCUCGGUGU	22	eca-miR-433	2.92E-45	-9.98
U140_3p	ACGCCCUUCCCCCCCUUCUUCA	22	eca-miR-1249	8.10E-24	-10.55
U29_3p	AUCACAUUGCCAGGGAUUUCC	21	eca-miR-23a	2.67E-09	-16.50

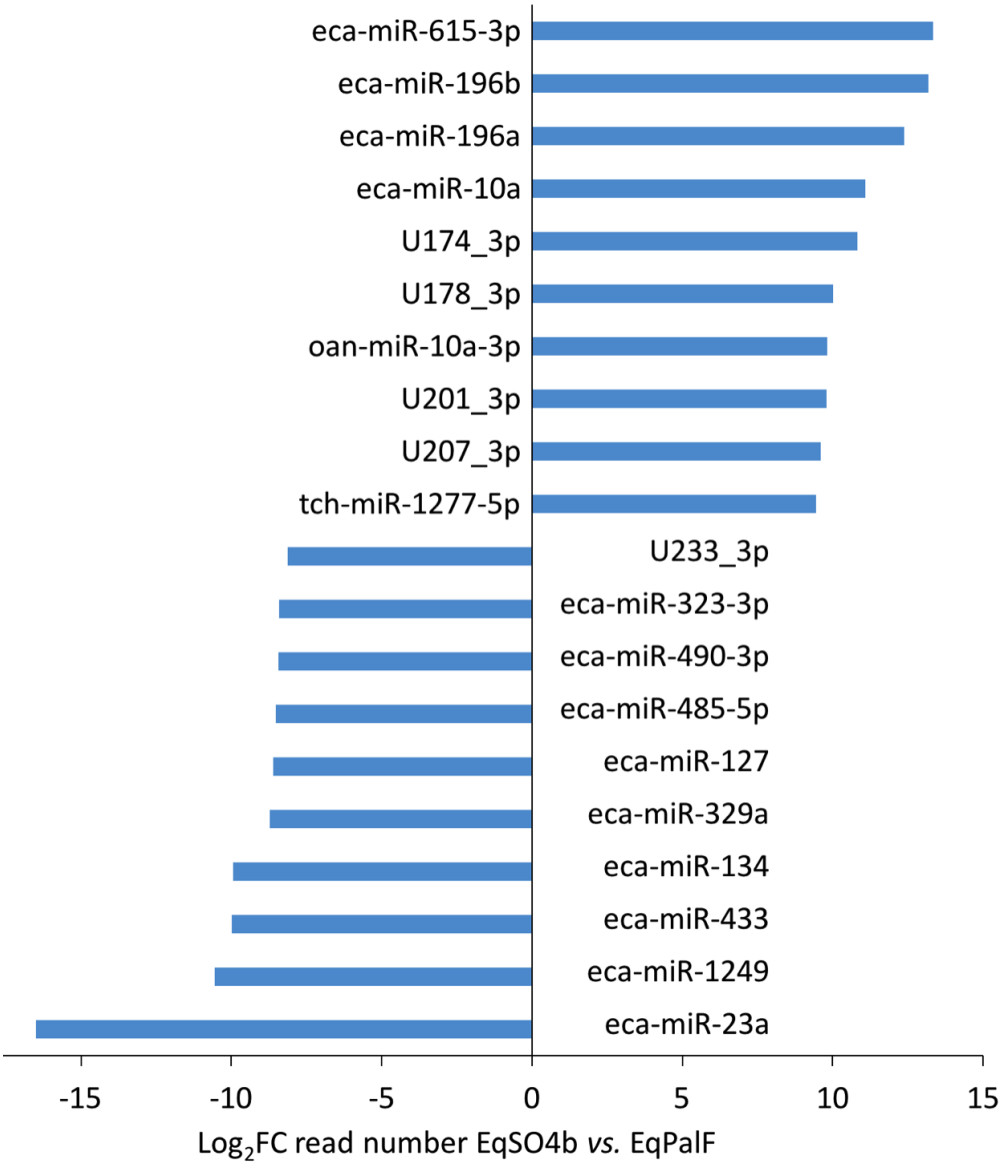


Figure 5.11. Top most differentially expressed miRNAs in EqSO4b vs. EqPalF (p adjusted <0.05) ranked by Log_2FC (ten most upregulated and ten most downregulated). When no match was found in miRBase 20.0, the miRNA maintained the internal project annotation (U_number_3p/5p). The x-axis denotes the Log_2FC between reads from EqSO4b and EqPalF: upregulated miRNAs in EqSO4b have values >0 and downregulated miRNAs have values <0

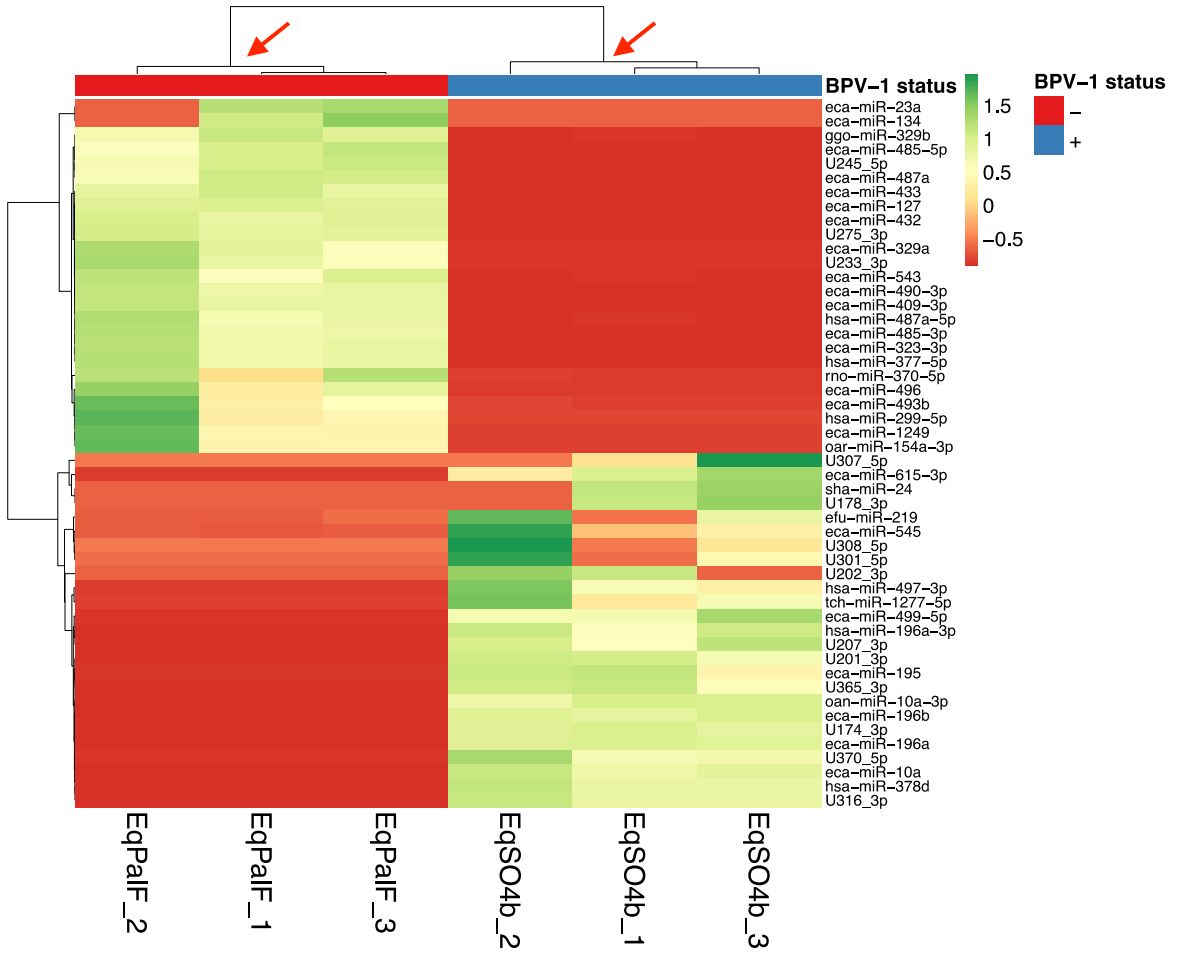


Figure 5.12. Heat map and hierarchical clustering of the top 50 most differentially expressed miRNAs in EqSO4b vs. EqPalF (Adjusted p<0.05) ranked by Log_2FC (25 most upregulated and 25 most downregulated). When no match for a miRNA was found in the database (miRBase) the internal identifier generated after harmonising the data, was used instead (U"number"_"3p/5p"). Hierarchical clustering was performed using correlation as clustering distance with average as the linkage criteria; the tree was ordered by the tightest cluster first. Each column represents one replicate (EqPalF_1,_2 and _3 for EqPalF and EqSO4b_1,_2 and _3 for EqSO4b). Note that through hierarchical clustering the replicates from each cell line are closer to each other than to the counterpart cell line, shown by the short branch lengths within both clusters (red arrows)

5.3.5. Validation of HTS data using real-time quantitative reverse-transcription polymerase chain reaction (qRT-PCR)

To validate the findings from the RNA seq data, qRT-PCR of a subset of miRNAs was carried out. Of the 1336 miRNAs (raw data), those that were found to be differentially expressed with adjusted $p<0.01$ in either of the pair-wise comparisons were ranked by absolute fold change. A more stringed adjusted p value was selected on this occasion as the aim of the qRT-PCR was validation of the HTS data. The top 60 differentially expressed miRNAs from each of the three pair-wise comparisons were merged into an Excel file. After 59 duplicates (i.e. miRNAs present on two or more lists) were removed, the remaining 121 miRNAs compared to the miRNAs in miRBase using in BLAST+ method described in Chapter 2. The criteria for perfect match was more restrictive than the criteria selected for discovery of novel miRNAs due to the importance of perfect match sequence inherent to the Taqman primers and probes. Only miRNAs with 100% identity and 100% coverage to a miRNA in the database within the list of equine miRNAs, were selected. Of the 44 miRNAs that had a perfect match to an equine miRNA, 37 had a commercially available Taqman® kit. Six miRNAs were selected from this list based on biological significance (i.e. associated with other tumours and/or linked to HPV-induced cancer) and most representative of changes between the three cell lines (i.e. number of reads, presence/absence in the three cell lines). For four miRNAs, data on fold changes between EqPalF and S6-2 obtained using microarray were also available (see Chapter 3; a summary of the microarray results for the miRNAs selected for validation is displayed in Table 5.5). For the other two miRNAs, microarray data were discarded as they were regarded as having expression below the background level. The list of six miRNAs selected for validation is displayed in Table 5.5.

Table 5.5. List of six miRNAs selected for qRT-PCR validation of HTS data

High throughput sequencing data								Microarray data (S6-2 *vs.* EqPalF)		Biological significance
ID and mature sequence	BLAST match*	EqSO4b *vs.* EqPalF		S6-2 *vs.* EqPalF		S6-2 *vs.* EqSO4b				
		Log_2FC	p-adj	Log_2FC	p-adj	Log2FC	p-adj	Log_2FC	p-adj	
U29_3p AUCACAUUGCCAGGGAUUUCC	eca-miR-23a	-16.50	2.67E-09	-0.16	1.00E+00	15.64	1.23E-08	-0.85	7.60E-07	Exclusive to non tumoural equine fibroblasts (EqPalF and S6-2). Downregulated in HPV induced cancer (Honegger *et al.*, 2015)
U9_3p UCGGAUCCGUCUGAGCUUGGCU	eca-miR-127	-8.62	4.43E-106	-4.44	5.52E-61	3.50	2.12E-14	-2.13	1.34E-08	Highly abundant. Downregulated in EqSO4b and S6-2. Reported downregulated in HPV-16 cervical cancer (Li *et al.*, 2011)
U97_5p UAGGUAGUUUCCUGUUGUUGGG	eca-miR-196b	13.19	1.20E-126	0.00	NA[#]	-12.30	4.00E-96	BBL**	BBL**	Exclusive to EqSO4b. Linked to tumorigenesis and invasion (Lu *et al.*, 2014)
U34_3p GAAUGUUGCUCGGUGAACCCCU	eca-miR-409-3p	-6.84	2.21E-76	-2.77	1.06E-26	3.40	2.94E-09	-1.17	5.36E-07	High read count. Present in all cells. Reported downregulated in HPV-16/18 samples (Lajer *et al.*, 2011)
U77_3p AAUAUAACACAGAUGGCCUGU	eca-miR-410	-6.01	1.15E-57	-2.37	4.06E-18	2.98	3.52E-03	BBL**	BBL**	High read count. Present in all cells. Reported downregulated in HVP-16/18 samples (Lajer *et al.*, 2011)
U140_3p ACGCCCUUCCCCCCCUUCUUCA	eca-miR-1249	-10.55	8.10E-24	-10.55	3.50E-22	-0.43	1.00E+00	0.10	1.53E-01	Exclusive to EqPalF and linked to TGF-b signalling (Vizoso *et al.*, 2015)

* match= 100% identity, 100% coverage

**BBL= below background level [#]NA= not applicable

The results of the qRT-PCR analysis of the six miRNAs selected, comparing the expression in EqSO4b and S62 to EqPalF (internal control) using $\Delta\Delta C_t$ method (for more details please see Chapter 2), and with RNU6 as normalizer, are displayed in Figure 5.13.

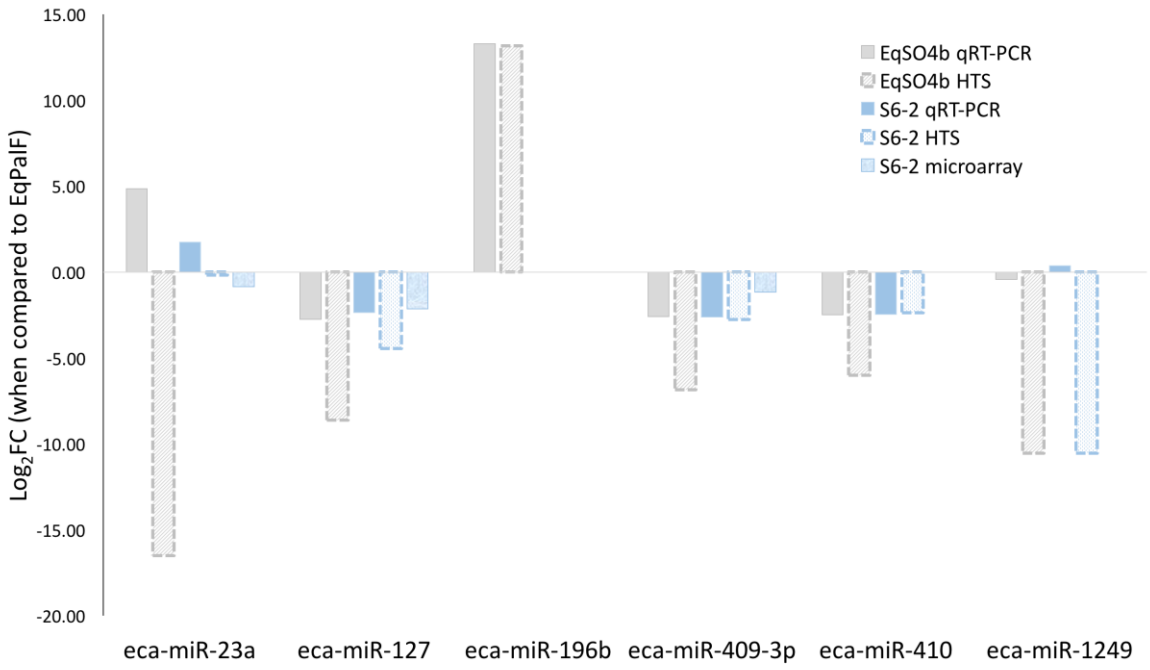


Figure 5.13. Histogram of expression of 6 miRNAs in EqSO4b and S6-2 selected for validation of HTS data using qRT-PCR. Results from HTS (dotted outline), qRT-PCR and microarray (only available for S6-2 cells) are displayed using EqPalF as control cell line and for qRT-PCR data is reported after normalisation to RNU6. Significance was adjusted to p-adj<0.05

qRT-PCR data for four miRNAs (eca-miR-127, eca-miR-196b, eca-miR-409-3p and eca-miR-410) successfully validated the HTS results. Of those, eca-miR-127, eca-miR-409-3p and eca-miR-410 were also downregulated in S6-2 cells compared to EqPalF from the previous microarray results (see Chapter 3). These three miRNAs have been reported to be downregulated in HPV-induced cancer: miR-127 downregulation has been found by microarray in tumours containing HPV-16 (Li *et al.*, 2011); miR-127, miR-409-3p and miR-410 have been found to be downregulated in tumours containing HPV-16/18 (Lajer *et al.*, 2011). Interestingly, miR-196b, a miRNA expressed in the same cluster as miR-196a, was not detected in EqPalF cells or S6-2 cells by any method and was highly expressed in EqSO4b cells as measured by both HTS and qRT-PCR. Discrepancies

in the levels of expression of miRNAs between HTS and qRT-PCR were found for eca-miR-23a and eca-miR-1249.

5.3.6. Differential expression of miRNAs commonly dysregulated in BPV-1 positive equine fibroblasts (S6-2 and EqSO4b) when compared to equine primary fibroblasts (EqPalF cells)

DESeq results

A total number of 201 miRNA candidates were commonly dysregulated in the cells containing BPV-1 genomes (EqSO4b and S6-2) when compared to EqPalF cells. In BPV-1 containing cells, 102 mature miRNAs candidates (51%) were found to be upregulated and 96 (48%) were found to be downregulated ($p<0.05$). Interestingly, out of the 201 commonly dysregulated miRNAs in BPV-1 positive cells, only three miRNAs (<1%) show disparity between the direction of the expression (either upregulated or downregulated) in EqSO4b and S6-2 cells. The rest (198 miRNAs, 99%) when they were upregulated in EqSO4b they were also upregulated in S6-2 cells. Level of expression of the 201 miRNAs had a positive and high correlation between the two cell lines containing BPV-1 genomes (Pearson's coefficient 0.91). Linear correlation of the level of expression (log_{10} transformed) of each of the 201 miRNAs is displayed in Figure 5.14. The top 25 most upregulated and top 25 most downregulated miRNAs in BPV-1 positive cells when compared to EqPalF is displayed in the heat map in Figure 5.15. The list of the top ten most up and ten most downregulated miRNAs is presented in Table 5.6.

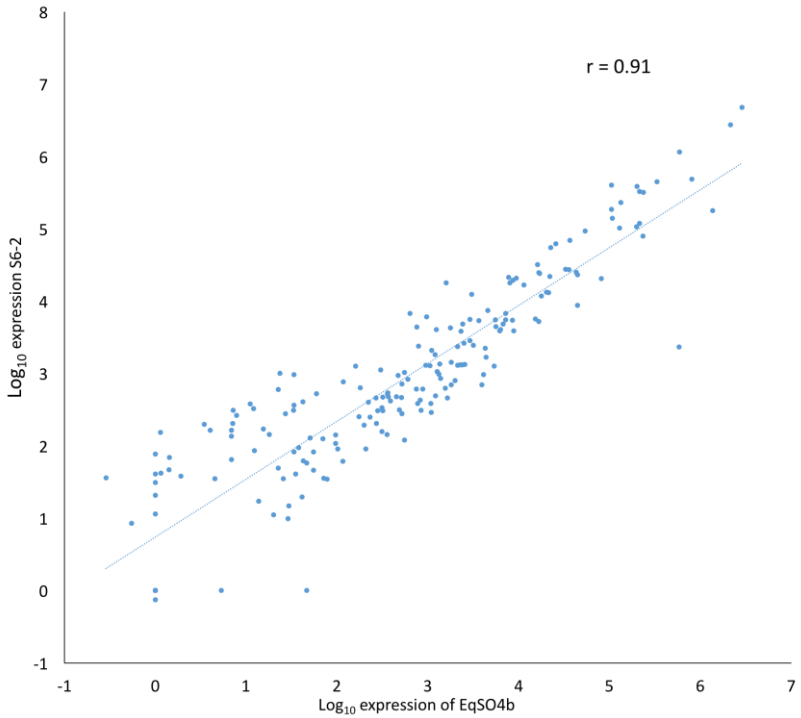


Figure 5.14. Correlation of expression of 201 miRNAs found dysregulated in BPV + cell lines when compared to control cell line. Each point in the graph corresponds to a miRNA with two coordinates: x-axis is the expression of this miRNA registered in S6-2 cells and y-axis is the expression of this miRNA registered in EqSO4b cells

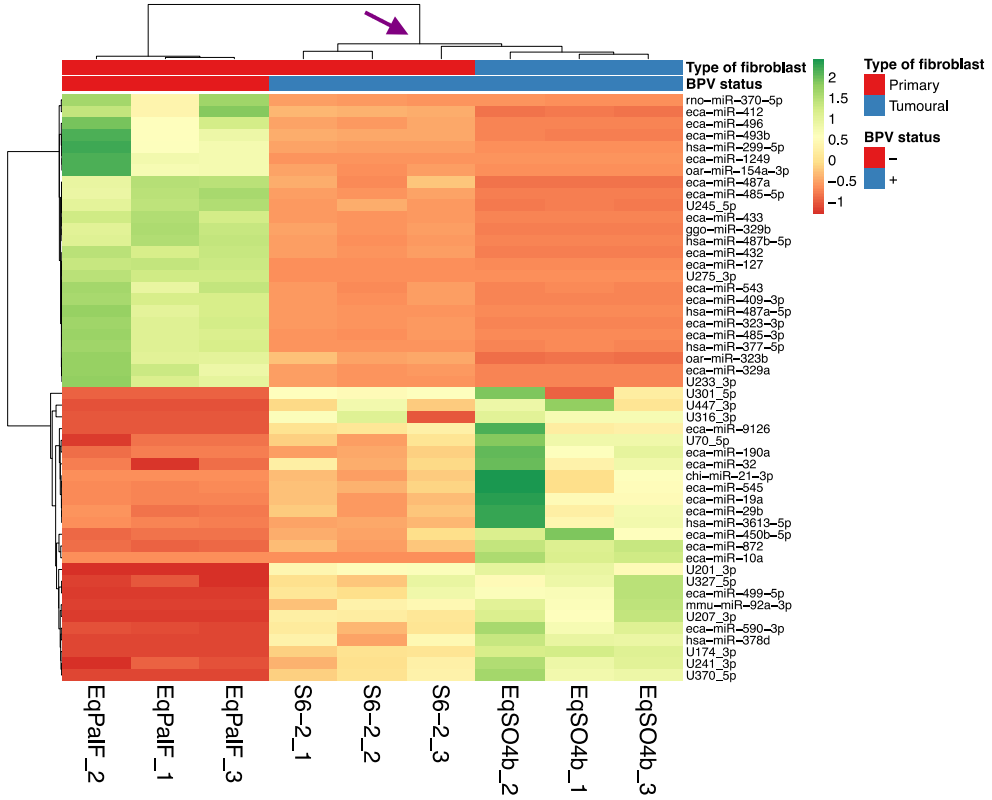


Figure 5.15. Heat map and hierarchical clustering of the top 50 most differentially expressed miRNAs in BPV-1 positive cells (S6-2 and EqSO4b) vs. EqPalF (p adjusted <0.05) ranked by log_2FC (25 most upregulated and 25 most downregulated). When no match for a miRNA was found in the database (miRBase 20.0) the internal identifier generated after harmonising the data, was used instead (Uxx-3p or 5p). Hierarchical clustering was performed using correlation as clustering distance between rows and columns, using the average as the linkage criteria and the tree was ordered by the tightest cluster first. Each column represents one replicate of each cell line. Note that through hierarchical clustering the replicates from each cell line are closer to each other than to the different cell line, and that the S6-2 replicates are closer EqSO4b than to EqPalF (see purple arrow)

Table 5.6. List of the top 10 most upregulated and most downregulated miRNA candidates in BPV positive cells (p<0.05). miRNAs have been ranked by Log_2FC in EqSO4b vs. EqPalF cells. miRNAs with annotation U"number"_"5p"/"3p" were miRNAs with no match on the database. ECA signifies equine chromosome

ID	Mature sequence	Annotation	ECA	Log_2FC EqSO4b *vs.* EqPalF	Log_2FC S6-2 *vs.* EqPalF
U7_5p	UACCCUGUAGAUCCGAAUUUGU	eca-miR-10a	11	11.09	2.39
U174_3p	CUAGAAUGAAGCUCCUAGAGG	U174_3p	5	10.82	9.95
U201_3p	UUGCCUGGGCCUCUGGAACCAU	U201_3p	1	9.79	9.76
U207_3p	UCAAGGAGAUCACAGUCUAGU	U207_3p	5	9.61	9.11
U143_3p	AUCAACAAACAUUUAUUGUGUGC	eca-miR-545	X	7.33	5.81
U316_3p	ACAAAAGUCUGCACAUUGGA	U316_3p	10	6.85	6.49
U322_5p	UUAAGACUUGCAGUGAUGUUU	eca-miR-499-5p	22	6.66	6.33
U359_3p	ACUGGACUUGGAGUCUGAAA	hsa-miR-378d	24	6.28	5.6
U301_5p	AGCCAGCUUGAGCCCACCUACAU	U301_5p	8	6.21	6.51
U370_5p	UCCGCGGAUUGAACCAGCCACG	U370_5p	18	6.04	5.31
U245_5p	CUGCUUUGCUCCCACUUUGUCCU	U245_5p	19	-7.67	-1.99
U184_5p	GUGGUUAUCCCUGCUGUGUUCG	hsa-miR-487a-5p	24	-7.8	-3.49
U240_3p	UGAGUAUUACAUGGCCAAUCUC	eca-miR-496	24	-7.8	-2.14
U233_3p	CACAAUACACGGUCGACCUCU	U233_3p	24	-8.14	-3.07
U130_3p	CACAUUACACGGUCGACCUCU	eca-miR-323-3p	24	-8.42	-2.87
U108_5p	AGAGGCUGGCCGUGAUGAAUUC	eca-miR-485-5p	24	-8.53	-2.43
U9_3p	UCGGAUCCGUCUGAGCUUGGCU	eca-miR-127	24	-8.62	-4.44
U214_3p	AACACACCUAGUUAACCUCUUU	eca-miR-329a	24	-8.72	-2.65
U175_3p	AUCAUGAUGGGCUCCUCGGUGU	eca-miR-433	24	-9.98	-3.03
U140_3p	ACGCCCUUCCCCCCCUUCUUCA	eca-miR-1249	28	-10.55	-10.55

The differences were explored in more detail between the two BPV+ cell lines. Out of 102 commonly upregulated miRNA candidates in BPV-1 positive cells, two thirds (66%, 67 miRNAs) showed no significant difference in expression when comparing EqSO4b and S6-2, whereas 20% (20 miRNAs) were significantly higher in EqSO4b cells. Amongst the upregulated miRNAs with significantly higher expression in EqSO4b there were several members of the miR-17~92 cluster (miR-17, 19a, 19b, 92a).

Out of the 96 commonly downregulated miRNA candidates in BPV-1 positive cells, half (50%, 48 miRNAs) were not found to be significantly different between the two cell lines. The remaining 48 miRNAs were most often downregulated in EqSO4b (43 miRNAs downregulated, 5 miRNAs upregulated). In the group of downregulated with significantly lower expression in EqSO4b there were multiple miRNAs that were encoded within the same cluster (genomic region of 3Kbp).

In total, 21 miRNAs were found to be encoded in seven clusters, interestingly in chromosome 24 (ECA24). ECA24 was found to be the chromosome with higher density of miRNAs and higher number of miRNAs in equine primary fibroblasts (please see Chapter 4). Details of clustered miRNAs are displayed in Table 5.7.

Table 5.7. List of 21 miRNAs found downregulated in BPV-1 positive cells that were significantly lower in tumoural cells and encoded in 7 clusters

Internal ID	Mature sequence	ECA	Annotation	Cluster
U9_3p	UCGGAUCCGUCUGAGCUUGGCU	24	eca-miR-127	1
U96_5p	UCUUGGAGUAGGUCAUUGGGUGG	24	eca-miR-432	
U175_3p	AUCAUGAUGGGCUCCUCGGUGU	24	eca-miR-433	
U105_5p	UGGUUUACCGUCCCACAUACAU	24	hsa-miR-299-5p	2
U133_3p	UAUGUAACAUGGUCCACUAAC	24	mml-miR-379-3p	
U122_3p	UAUGUAAUAUGGUCCACGUCU	24	eca-miR-380	
U40_3p	UAUGUAACACGGUCCACUAAC	24	mml-miR-411-3p	
U130_3p	CACAUUACACGGUCGACCUCU	24	eca-miR-323-3p	3
U233_3p	CACAAUACACGGUCGACCUCU	24	U233_3p	
U176_3p	AACGAACCUGGUUAACCUCU	24	ggo-miR-329b	
U146_3p	AAACAAACAUGGUGCACUUCUU	24	eca-miR-495	4
U180_3p	AAACAUUCGCGGUGCACUUCUU	24	eca-miR-543	
U148_5p	GAAGUUGUUCGUGGUGGAUUCG	24	eca-miR-382	5
U108_3p	GUCAUACACGGCUCUCCUCUCU	24	eca-miR-485-3p	
U108_5p	AGAGGCUGGCCGUGAUGAAUUC	24	eca-miR-485-5p	
U165_5p	AGAGGUCUUCCAUGAUGCAUUCG	24	oar-miR-154b-5p	6
U196_5p	AGAGGUUGCCCUUGGUGAAUUC	24	hsa-miR-377-5p	
U240_3p	UGAGUAUUACAUGGCCAAUCUC	24	eca-miR-496	
U34_3p	GAAUGUUGCUCGGUGAACCCCU	24	eca-miR-409-3p	7
U194_5p	UGGUCGACCAGUUGGAAAGUAAU	24	eca-miR-412	
U203_5p	AAAGGAUUCUGCUGUCGGUCCCACU	24	hsa-miR-541-5p	

Other miRNAs of interest in the list of downregulated miRNAs in BPV-1 positive cells were miR-181a, miR-181b and miR-204b. As previously mentioned, miR-181a (U4_5p) and miR-181b (U44_5p) have been reported to be downregulated in HPV-16/-18 positive cell cultures with the use of microarray and qRT-PCR (Lajer *et al.*, 2011; Wald *et al.*, 2011) and they are encoded in a region of ECA25 recently reported to be associated with the development of ES (Jandova *et al.*, 2012). miR-204b (U_72_5p) has been reported to be downregulated in cervical cancer using HTS (Witten *et al.*, 2010) and it is encoded in equine fibroblasts very close to one of the regions recently identified as deleted and linked to ES (Bugno-Poniewierska *et al.*, 2016).

Comparison of HTS data with microarray and miRNAs previously reported to be dysregulated in HPV

When the 201 miRNA candidates found dysregulated in BPV-1 positive cells compared to EqPalF cells using HTS were compared to previous data from the microarray (comparing S6-2 cells to EqPalF cells, see Chapter 3), a total of 52 miRNAs had information from both technologies. The remaining 149 were either not interrogated by microarray or showed levels below background detection. Just over one third of the miRNAs where data was available from both methods (34%, 17 miRNAs) had results that agreed in HTS and microarray. Three of those were also assessed using SYBR green - qRT-PCR as part of the validation of microarray data (miR-181a and miR-15b; see Chapter 3, section 3.3.3.) or Taqman® – qRT-PCR as part of the validation of HTS data (miR-127, see previous section 5.3.5). miR-181a and miR-127 were shown to be downregulated in BPV-1 positive cells, in agreement with both HTS and microarray, whereas miR-15b was found to be upregulated using qRT-PCR and downregulated in HTS and microarray.

When the 17 miRNAs (where information was available for both HTS and microarray), were compared to a database generated for the current study from 13 experimental papers on HPV related cancer and/or high-risk HPV cell cultures both from *in vitro* and *ex vivo* studies and four review papers (Wang *et al.*, 2008; Pereira *et al.*, 2010; Witten *et al.*, 2010; Greco *et al.*, 2011; Li *et al.*, 2011; McBee *et al.*, 2011; Wald *et al.*, 2011; Yeung *et al.*, 2011; Zheng & Wang, 2011; Lajer *et al.*, 2012; Gocze *et al.*, 2013; Gómez-Gómez *et al.*, 2013; Sharma *et*

al., 2014; Wang *et al.*, 2014b; Ben *et al.*, 2015; Honegger *et al.*, 2015; Wang *et al.*, 2016), nine of those were also found dysregulated in HPV studies, of which six were dysregulated in HPV and BPV (U11_5p, U30_5p, U4_5p, U44_5p, U72_5p and U9_3p).Three out of the six dysregulated in HPV were found to be encoded in genomic regions previously linked to equine sarcoids (U4_5p, U44_5p and U72_5p) (Jandova *et al.*, 2012; Bugno-Poniewierska *et al.*, 2016). Details on the 17 miRNAs are displayed in Table 5.8.

Table 5.8. List of 17 miRNAs found dysregulated in BPV-1 + cells in the current study using HTS and microarray. Out of the 17, 9 were reported to be dysregulated in the literature in HPV related disease/cancer and six were dysregulated in the same way as reported in HPV. It is worth noting that some studies on microRNA in HPV show disparity of results. miRNAs highlighted in grey are encoded in regions of the equine genome reported to be linked to equine sarcoids

ID	Mature sequence	ECA	Annotation	Result in HTS and microarray	Results database (16 studies)
U105_5p	UGGUUUACCGUCCCACAUACAU	24	hsa-miR-299-5p	downregulated in BPV-1 +	Not on HPV database
U108_3p	GUCAUACACGGCUCUCCUCUCU	24	eca-miR-485-3p	downregulated in BPV-1 +	Not on HPV database
U11_5p	CAACGGAAUCCCAAAAGCAGCUG	16	eca-miR-191a	downregulated in BPV-1 +	Low in CC and HPV+ cell line using microarray (Wang et al., 2008)
U148_5p	GAAGUUGUUCGUGGUGGAUUCG	24	eca-miR-382	downregulated in BPV-1 +	Not on HPV database
U30_5p	UGAGGGGCAGAGAGCGAGACUUU	11	eca-miR-423-5p	downregulated in BPV-1 +	Low in HPV-16/18 cell line using microarray and qRT-PCR (Lajer et al., 2012)
U34_3p	GAAUGUUGCUCGGUGAACCCCU	24	eca-miR-409-3p	downregulated in BPV-1 +	Not on HPV database
U4_5p	AACAUUCAACGCUGUCGGUGAGU	25	eca-miR-181a	downregulated in BPV-1 +	High in HPV + monolayer culture but low in HPV + raft tissues using microarray (Wang et al., 2008), low in HPV-16 cell line using microarray and qRT-PCR (Wald et al., 2011), low in HPV-16/18 cell line using microarray and qRT-PCR (Lajer *et al.*, 2011). Encoded in ECA25, region associated with equine sarcoids (Jandova et al., 2012)
U44_5p	AACAUUCAUUGCUGUCGGUGGGU	25	eca-miR-181b	downregulated in BPV-1 +	High in HPV + monolayer culture but low in HPV + raft tissues using microarray (Wang et al., 2008); low in HPV-16 cell line using microarray and qRT-PCR (Wald et al., 2011), low in HPV-16/18 cell line using microarray and qRT-PCR (Lajer et al., 2012). Encoded in ECA25, region associated with equine sarcoids (Jandova et al., 2012)
U58_5p	UAGCAGCACAUCAUGGUUUACA	5	eca-miR-15b	downregulated in BPV-1 +	High in CC using microarray (Wang et al., 2008); high in HPV-16 SCC using microarray and qRT-PCR (Li et al., 2011) and high in HPV-16/18 cell line (Lajer et al., 2012)
U64_3p	UAACAGUCUACAGCCAUGGUCG	11	eca-miR-132	downregulated in BPV-1 +	High in CC HPV-16/18 using microarray (Pereira et al., 2010)
U72_5p	UUCCCUUUGUCAUCCUAUGCCU	23	eca-miR-204b	downregulated in BPV-1 +	Low in cancer tissue RNA seq (Witten et al., 2010). Encoded in ECA 23 very close to the region recently identified as deleted and linked to ES (Bugno-Poniewierska *et al.*, 2016)

(Continuation Table 5.8)

U73_5p	UGGGUCUUUGCGGGCGAGAUGA	11	eca-miR-193a-5p	downregulated in BPV-1 +	Not on HPV database
U9_3p	UCGGAUCCGUCUGAGCUUGGCU	24	eca-miR-127	downregulated in BPV-1 +	Low in CC lesions using microarray (Wang et al., 2008) and microarray and qRT-PCR (Li et al., 2011)
U96_5p	UCUUGGAGUAGGUCAUUGGGUGG	24	eca-miR-432	downregulated in BPV-1 +	Not on HPV database
U100_3p	UAGCACCAUUUGAAAUCAGUGUU	4	eca-miR-29b	upregulated in BPV-1 +	Not on HPV database
U158_3p	UGUGACAGAUUGAUAACUGAAA	X	eca-miR-542-3p	upregulated in BPV-1 +	Not on HPV database
U76_5p	AACCCGUAGAUCCGAUCUUGUG	26	eca-miR-99a	upregulated in BPV-1 +	Low in CC using microarray (Wang et al., 2008; Pereira et al., 2010), using microarray and qRT-PCR (Li et al., 2011), low in HPV-16/18 cell line using microarray and qRT-PCR (Lajer *et al.*, 2012)

5.3.7. miRNAs dysregulated in tumoural equine fibroblasts but not in BPV-1 transformed cells

When compared with control cells (EqPalF), a total number of 329 miRNAs were found to be dysregulated (either up or down) in EqSO4b cells (see section 5.3.4). Of those, 201 were dysregulated in both EqSO4b and S6-2 cells when compared to EqPalF. The rest, 128 miRNA candidates were dysregulated in EqSO4b but not in S6-2 when compared to EqPalF. Out of 128 miRNA candidates, 63 (49%) miRNAs were downregulated and 65 miRNAs were upregulated (51%) in EqSO4b cells when compared to EqPalF cells.

In the list of downregulated miRNAs, U22_5p (eca-miR-145) and U24_3p (eca-miR-23b) were of particular interest. These miRNAs have been reported to be linked to HR-HPV cancer: miR-145 has been reported to be downregulated *in vitro* (Wang *et al.*, 2008) and *ex vivo* (Pereira *et al.*, 2010; Li *et al.*, 2011) due to HR-HPV E6 mediated downregulation of p53; and miR-23b has been reported to be downregulated via HPV-16 E5 and E6 oncoproteins (Greco *et al.*, 2011; Yeung *et al.*, 2011; Honegger *et al.*, 2015). Amongst the 65 upregulated miRNAs, several have been reported to be upregulated in HR-HPV-induced cancer in *in vitro* and *ex vivo* studies. Of particular interest are U16_5p (miR-16) and U120_5p (miR-196a), which were upregulated in HR-HPV cancer via E7 oncoprotein (Wang *et al.*, 2008; Pereira *et al.*, 2010; Li *et al.*, 2011; Lajer *et al.*, 2012; Gocze *et al.*, 2013). A list of the miRNAs dysregulated in EqSO4b in the same direction as in studies on HR-HPV related cancer and the references for these studies, are displayed in Table 5.9.

Table 5.9. List of miRNAs that were found dysregulated in tumoural equine fibroblasts and HPV-induced cancer

Internal ID	Annotation	Expression in EqSo4b (*vs.*EqPalF)	Studies showing same direction of expression change of this miRNA in HPV-16/-18 induced cancer
U14_5p	eca-miR-125a-5p	Downregulated	Wang *et al.* (2008)
U121_3p	eca-miR-133a	Downregulated	Wang *et al.* (2008)
U22_5p	eca-miR-145	Downregulated	Wang *et al.* (2008); Pereira *et al.* (2010); Li *et al.* (2011)
U24_3p	eca-miR-23b	Downregulated	Greco *et al.* (2011); Yeung *et al.* (2011); Honegger *et al.* (2015)
U5_5p	hsa-miR-27b-5p	Downregulated	Wang *et al.* (2008)
U225_5p	eca-miR-539	Downregulated	Greco *et al.* (2011)
U18_5p	eca-let-7a	Upregulated	Wang *et al.* (2008)
U92_5p	eca-miR-106a	Upregulated	Wang *et al.* (2008); Pereira *et al.* (2010); Li *et al.* (2011); Lajer *et al.* (2012)
U7_3p	oan-miR-10a-3p	Upregulated	Wang *et al.* (2008); Pereira *et al.* (2010)
U126_3p	eca-miR-130b	Upregulated	Lajer *et al.* (2012)
U227_3p	bta-miR-142-3p	Upregulated	Witten *et al.* (2010)
U227_5p	eca-miR-142-5p	Upregulated	Pereira *et al.* (2010)
U78_5p	eca-miR-146b-5p	Upregulated	Wang *et al.* (2008); Lajer *et al.* (2012)
U28_3p	eca-miR-148a	Upregulated	Wang *et al.* (2008); Pereira *et al.* (2010)
U16_3p	gga-miR-16-1-3p	Upregulated	Wang *et al.* (2008); Pereira *et al.* (2010); Li *et al.* (2011); McBee *et al.* (2011); Lajer *et al.* (2012)
U173_5p	eca-miR-192	Upregulated	McBee *et al.* (2011)
U120_5p	eca-miR-196a	Upregulated	Pereira *et al.* (2010); Gocze *et al.* (2013)
U115_3p	mml-miR-20a-3p	Upregulated	Wang *et al.* (2008); Li *et al.* (2011)
U17_5p	mmu-miR-221-5p	Upregulated	Wang *et al.* (2008); Li *et al.* (2011); Gocze *et al.* (2013)
U229_5p	eca-miR-7	Upregulated	Li *et al.* (2011)
U55_5p	eca-miR-93	Upregulated	Wang *et al.* (2008); Li *et al.* (2011); Lajer *et al.* (2012)

5.3.8. Functional analysis of miRNAs dysregulated in the *in vitro* equine sarcoid model

Functional analysis on the miRNAs associated with the presence of BPV + and up or downregulated in S6-2 and EqSO4b cells was carried out using Ingenuity Pathway Analysis (IPA). For this purpose the sequences of the miRNAs found to be dysregulated were matched with any human or mouse miRNAs in miRBase 20.0 to obtain a numerical annotation (MIMAxxxxx) of the homologous miRNA in those species with a tolerance of two mismatches.

Out of 329 miRNAs dysregulated in EqSO4b cells and 251 in S6-2 cells compared to EqPalF, the software recognised 173 and 103 miRNAs respectively. The number of miRNAs mapped to the IPA database did not retrieve any information on canonical pathways or upstream analysis. However, when using core analysis, the top networks associated with the dysregulated miRNAs in tumoural cells were: reproductive system disease (Cancer), reproductive system disease (cellular development) and connective tissue disorders and cellular development-growth and proliferation (cancer). The top molecular and cellular functions were cellular development (36 molecules), cellular growth and proliferation (33 molecules), cell cycle (16 molecules), cellular movement (20 molecules) and DNA repair (eight molecules). Similarly, the top networks associated with the dysregulated miRNAs in BPV-1 transformed cells were: reproductive system disease (cancer), reproductive and connective tissue disorders (organismal injury and abnormalities) and connective tissue disorders (cancer). The top molecular and cellular functions were cellular development (32 molecules), cellular growth and proliferation (34 molecules), cell cycle (14 molecules), cellular movement (17 molecules) and cell death and survival (23 molecules). The list of diseases, functions and networks from both comparisons (EqSO4b *vs.* EqPalF and S6-2 *vs.* EqPalF) are displayed in Tables 5.10 and 5.11.

Table 5.10. Diseases, functions and networks obtained from pathway analysis from the miRNAs dysregulated in EqSO4b vs EqPalF (using Ingenuity Pathway Analysis software)

Top diseases and biofunctions			
Diseases and disorders	p-value range		Number of molecules
Organismal injury and abnormalities	4.70E-02	6.92E-44	94
Reproductive system disease	3.41E-02	6.92E-44	62
Cancer	4.70E-02	7.61E-32	70
Gastrointestinal disease	3.26E-02	2.70E-27	48
Molecular and cellular functions	p-value range		Number of molecules
Cellular development	4.72E-02	1.92E-06	36
Cellular growth and proliferation	4.72E-02	1.92E-06	33
Cell cycle	4.70E-02	2.45E-05	16
Cellular movement	3.67E-02	8.69E-05	20
DNA replication, recombination and repair	4.33E-02	6.19E-04	8
Physiological system development and function	p-value range		Number of molecules
Organismal development	6.48E-03	1.23E-07	7
Connective tissue development and function	1.59E-02	2.45E-05	6

Top networks	Score
Organismal injury and abnormalities, reproductive system disease, cancer	60
Organismal injury and abnormalities, reproductive system disease, cellular development	35
Developmental disorder, hereditary disorder, organismal functions	28
Cancer, connective tissue disorders, organismal injury and abnormalities	24
Cellular development, cellular growth and proliferation, cancer	24

Table 5.11. Diseases, functions and networks obtained from pathway analysis from the miRNAs dysregulated in S6-2 vs EqPalF (using Ingenuity Pathway Analysis software)

Top diseases and biofunctions			
Diseases and disorders	p-value range	Number of molecules	
Organismal injury and abnormalities	4.85E-02	7.29E-41	84
Reproductive system disease	1.71E-02	7.29E-02	56
Cancer	4.85E-02	4.38E-30	66
Gastrointestinal disease	4.78E-02	2.61E-29	55
Molecular and cellular functions	p-value range	Number of molecules	
Cellular development	4.75E-02	5.18E-08	32
Cellular growth and proliferation	4.75E-02	5.18E-08	34
Cell cycle	3.82E-02	1.06E-05	14
Cellular movement	2.98E-02	3.15E-05	17
Cell death and survival	4.65E-02	1.10E-04	23
Physiological system development and function	p-value range	Number of molecules	
Organismal development	3.77E-02	4.22E-08	6
Connective tissue development and function	2.56E-02	1.06E-05	5

Top networks	Score
Organismal injury and abnormalities, reproductive system disease, cancer	64
Organismal injury and abnormalities, reproductive system disease, connective tissue	25
Organismal injury and abnormalities, reproductive system disease, developmental disorders	25
Cancer, connective tissue dissorders, organismal injury and abnormalities	24
Cancer, organismal injury and abnormalities, gastrointestinal disease	24

As only the miRNAs with homologous human or mice could be used for functional analysis, and only miRNAs previously added to the database of the software are recognised, the list of top ten miRNAs most up and downregulated identified by the IPA software was dissimilar to the lists displayed in Tables 5.3. and Table 5.4. The top ten miRNAs most up and downregulated from the list uploaded in IPA are listed in Table 5.12 and Table 5.13. Of those, it is worth noting that miR-151-3p/5p, miR-499-5p and miR-92a-3p were found in the top most upregulated in both comparisons and that miR-127-3p, miR-487a-5p and miR-485-3p were found in the top 10 most downregulated in both comparisons. Other miRNAs of interest are miR-196a-5p, mir-23 precursor and miR-485 as they were found to be involved in multiple networks: organismal injury abnormalities-reproductive system disease-cancer and organismal injury and abnormalities-reproductive system disease-cellular development and cancer, connective tissue disorders, organismal injury and abnormalities (mir-23 and mature forms); and organismal injury abnormalities-reproductive system disease-cancer and organismal injury and abnormalities-reproductive system disease-cellular development (miR-196a-5p and miR-485).

Other miRNAs not in the top most up or downregulated of the software IPA but found of interest were miR-181, miR-29b, miR-10b and miR-127 as they are in the list of molecules that are dysregulated in the networks associated with developmental disorder-hereditary disorder-organismal functions and cancer-connective tissue disorders-organismal injury and abnormalities.

Table 5.12. List of top ten up and downregulated miRNAs identified during pathway analysis of miRNAs found dysregulated between EqSO4b cells and EqPalF cells. The software recognised 173 out of the 329 miRNAs found differentially expressed (adj p<0.05)

Upregulated
miR-196a-5p (and other miRNAs w/seed AGGUAGU)
miR-151-3p (and other miRNAs w/seed UAGACUG)
miR-151-5p (and other miRNAs w/seed CGAGGAG)
miR-196a-3p (miRNAs w/seed GGCAACA)
miR-499-5p (and other miRNAs w/seed UAAGACU)
miR-378a-3p (and other miRNAs w/seed CUGGACU)
miR-497-3p (and other miRNAs w/seed AAACCAC)
miR-92a-3p (and other miRNAs w/seed AUUGCAC)
miR-16-5p (and other miRNAs w/seed AGCAGCA)
miR-3613-5p (miRNAs w/seed GUUGUAC)
Downregulated
miR-23a-3p (and other miRNAs w/seed UCACAUU)
miR-1249-3p (and other miRNAs w/seed CGCCCUU)
miR-433-3p (miRNAs w/seed UCAUGAU)
miR-3118 (and other miRNAs w/seed GUGACUG)
miR-127-3p (miRNAs w/seed CGGAUCC)
miR-485-5p (and other miRNAs w/seed GAGGCUG)
miR-323-3p (and other miRNAs w/seed ACAUUAC)
miR-503-3p (and other miRNAs w/seed GAGUAUU)
miR-487a-5p (and other miRNAs w/seed UGGUUAU)
miR-485-3p (and other miRNAs w/seed UCAUACA)

Table 5.13. List of top ten up and downregulated miRNAs identified during pathway analysis of miRNAs found dysregulated between S6-2 cells and EqPalF cells. The software recognised 103 out of the 251 miRNAs found differentially expressed (p<0.05)

Upregulated
miR-151-3p (and other miRNAs w/seed UAGACUG)
miR-151-5p (and other miRNAs w/seed CGAGGAG)
miR-802-5p (miRNAs w/seed CAGUAAC)
miR-499-5p (and other miRNAs w/seed UAAGACU)
miR-378a-3p (and other miRNAs w/seed CUGGACU)
let-7a-3p (and other miRNAs w/seed UAUACAA)
miR-92a-3p (and other miRNAs w/seed AUUGCAC)
miR-1-3p (and other miRNAs w/seed GGAAUGU)
miR-590-3p (miRNAs w/seed AAUUUUA)
miR-143-5p (and other miRNAs w/seed GUGCAGU)
Downregulated
miR-1249-3p (and other miRNAs w/seed CGCCCUU)
miR-146a-5p (and other miRNAs w/seed GAGAACU)
miR-184 (and other miRNAs w/seed GGACGGA)
miR-122-5p (miRNAs w/seed GGAGUGU)
miR-224-5p (miRNAs w/seed AAGUCAC)
miR-127-3p (miRNAs w/seed CGGAUCC)
miR-370-5p (and other miRNAs w/seed AGGUCAC)
miR-708-5p (and other miRNAs w/seed AGGAGCU)
miR-485-3p (and other miRNAs w/seed UCAUACA)
miR-487a-5p (and other miRNAs w/seed UGGUUAU)

5.3.9. Biomarker study

Using the list of differentially expressed miRNAs between S6-2 cells *vs.* EqPalF and EqSO4b *vs.* EqPalF, the linear discriminant analysis (LDA) effect size tool (LEfSe, Galaxy, https://huttenhower.sph.harvard.edu/galaxy), identified 59 potential biomarkers for BPV-1 (Table 5.14) and 47 potential biomarkers for tumoural cells (tumoural *vs.* primary, Table 5.15). Examples of the relative expression levels of some miRNAs as biomarkers for BPV-1 and miRNAs as biomarkers for tumoural phenotype in the three cell lines of the ES *in vitro* model are presented in Figure 5.16. and 5.17. respectively. The mature sequence or the annotation name/number from miRBase is not relevant for this type of analysis and therefore the biomarker study conducted on LEfSe highlighted multiple novel miRNAs as suitable biomarker candidates that could not have been included in the functional analysis.

Table 5.14. List of miRNAs selected as biomarkers for BPV-1 presence

Internal ID	Annotation	Log2FC EqSO4b vs EqPalF	Log2FC S6-2 vs EqPalF	Direction HTS EqSO4b vs EqPalF	Direction HTS S6-2 vs EqPalF	Direction HTS S6-2 vs EqSO4b
U1_3p	mmu-miR-21-3p	4.35	2.70	Up	Up	not DE
U1_5p	eca-miR-21	2.05	2.07	Up	Up	not DE
U10_3p	eca-miR-92a	-1.52	-1.73	Down	Down	Down
U105_5p	hsa-miR-299-5p	-7.59	-2.88	Down	Down	Up
U107_5p	U107_5p	3.06	3.27	Up	Up	Up
U108_3p	eca-miR-485-3p	-7.62	-3.50	Down	Down	Up
U11_5p	eca-miR-191a	-1.39	-1.30	Down	Down	Up
U111_3p	eca-miR-454	2.07	2.42	Up	Up	not DE
U114_3p	eca-miR-483	-3.99	-1.46	Down	Down	Up
U117_3p	eca-miR-671-3p	-1.95	-2.20	Down	Down	not DE
U118_5p	U118_5p	-1.74	-1.57	Down	Down	not DE
U122_3p	eca-miR-380	-4.83	-1.43	Down	Down	Up
U125_5p	eca-miR-149	-1.76	-2.60	Down	Down	not DE
U129_3p	eca-miR-7177b	-1.34	-2.56	Down	Down	Down
U13_3p	eca-miR-222	-0.86	-1.74	Down	Down	Down
U130_3p	eca-miR-323-3p	-8.42	-2.87	Down	Down	Up
U16_5p	eca-miR-16	1.48	1.19	Up	Up	not DE
U20_5p	bta-miR-6119-5p	2.27	1.97	Up	Up	not DE
U23_5p	eca-let-7f	1.73	1.95	Up	Up	Up
U27_3p	eca-miR-27a	1.98	1.83	Up	Up	Down
U3_5p	eca-miR-10b	2.45	-1.21	Up	Down	Down
U30_5p	eca-miR-423-5p	-1.87	-1.41	Down	Down	Up
U34_3p	eca-miR-409-3p	-6.84	-2.77	Down	Down	Up
U36_3p	hsa-let-7d-3p	-1.88	-1.29	Down	Down	not DE
U37_5p	eca-miR-30e	2.72	2.77	Up	Up	Up
U4_3p	efu-miR-181f	-3.03	-2.25	Down	Down	not DE
U4_5p	eca-miR-181a	-2.57	-1.36	Down	Down	Up
U40_3p	mml-miR-411-3p	-5.70	-1.87	Down	Down	Up
U41_3p	eca-miR-101	3.19	1.90	Up	Up	Down

U43_3p	eca-miR-193b	-1.11	-1.01	Down	Down	not DE
U44_5p	eca-miR-181b	-2.00	-1.27	Down	Down	Up
U47_5p	eca-miR-340-5p	3.26	2.28	Up	Up	not DE
U51_5p	eca-let-7g	1.59	1.79	Up	Up	Up
U53_5p	bta-miR-6529a	-2.38	-1.07	Down	Down	Up
U54_3p	eca-miR-301a	3.29	3.03	Up	Up	Down
U56_3p	eca-miR-532-3p	-1.27	-1.28	Down	Down	not DE
U58_5p	eca-miR-15b	-1.85	-1.99	Down	Down	not DE
U59_5p	eca-miR-1388	2.59	2.35	Up	Up	not DE
U6_3p	eca-miR-199a-3p	1.50	1.48	Up	Up	not DE
U6_5p	eca-miR-199b-5p	1.96	1.86	Up	Up	not DE
U60_5p	eca-miR-15a	2.70	1.96	Up	Up	not DE
U61_3p	eca-miR-197	-1.85	-1.54	Down	Down	not DE
U64_3p	eca-miR-132	-4.21	-1.79	Down	Down	not DE
U66_3p	eca-miR-342-3p	-1.17	-1.11	Down	Down	not DE
U67_5p	eca-miR-98	2.50	2.33	Up	Up	not DE
U68_5p	bta-miR-1260b	-3.48	-1.62	Down	Down	Up
U72_5p	eca-miR-204b	-2.38	-1.81	Down	Down	Up
U73_3p	eca-miR-193a-3p	3.20	2.46	Up	Up	not DE
U73_5p	eca-miR-193a-5p	-2.32	-1.17	Down	Down	Up
U74_5p	hsa-miR-128-1-5p	-1.76	-2.14	Down	Down	not DE
U77_3p	eca-miR-410	-6.01	-2.37	Down	Down	Up
U79_5p	eca-miR-486-5p	-3.75	-2.88	Down	Down	Up
U80_3p	ssc-miR-374a-3p	2.83	2.34	Up	Up	Down
U80_5p	eca-miR-374a	1.94	1.92	Up	Up	not DE
U82_5p	eca-miR-450b-5p	3.71	2.52	Up	Up	not DE
U86_3p	mml-miR-106b-3p	-1.36	-1.60	Down	Down	Down
U89_3p	eca-miR-3200	-1.64	-1.73	Down	Down	not DE
U9_3p	eca-miR-127	-8.62	-4.44	Down	Down	Up
U96_5p	eca-miR-432	-7.19	-2.49	Down	Down	Up

Table 5.15. List of miRNAs selected as biomarkers for tumoural transformation (tumour vs primary cells)

Internal ID	Annotation	Log2FC EqSO4b vs EqPalF	Log2FC S6-2 vs EqPalF	Direction HTS EqSO4b vs EqPalF	Direction HTS S6-2 vs EqPalF	Direction HTS S6-2 vs EqSO4b
U100_3p	eca-miR-29b	4.07	2.67	Up	Up	Down
U102_3p	eca-miR-381	-3.32	0.16	Down	not DE	Up
U103_5p	eca-miR-450c	3.22	2.14	Up	Up	not DE
U109_3p	hsa-miR-136-3p	-2.47	1.06	Down	not DE	Up
U113_3p	eca-miR-889	-4.90	-0.51	Down	not DE	Up
U119_3p	eca-miR-487b	-4.72	-0.34	Down	not DE	Up
U120_5p	eca-miR-196a	12.37	0.00	Up	not DE	Down
U121_3p	eca-miR-133a	-0.93	0.57	Down	not DE	not DE
U14_5p	eca-miR-125a-5p	-0.77	-0.52	Down	not DE	Up
U201_3p	U201_3p	9.79	9.76	Up	Up	not DE
U22_5p	eca-miR-145	-2.62	-0.74	Down	not DE	Up
U223_5p	eca-miR-490-5p	-5.14	-1.79	Down	not DE	not DE
U24_3p	eca-miR-23b	-1.69	-0.33	Down	not DE	not DE
U26_5p	eca-miR-186	1.18	1.25	Up	Up	Up
U275_3p	U275_3p	-7.15	-6.95	Down	Down	not DE
U28_3p	eca-miR-148a	3.23	1.08	Up	not DE	Down
U316_3p	U316_3p	6.85	6.49	Up	Up	not DE
U32_3p	eca-miR-29a	1.16	1.52	Up	Up	not DE

U359_3p	hsa-miR-378d	6.28	5.60	Up	Up	not DE
U370_5p	U370_5p	6.04	5.31	Up	Up	not DE
U371_5p	eca-miR-9126	5.63	5.08	Up	Up	not DE
U40_5p	eca-miR-411	-3.77	0.23	Down	not DE	Up
U429_5p	U429_5p	5.51	3.71	Up	not DE	not DE
U435_3p	mmu-miR-92a-3p	5.38	4.82	Up	Up	not DE
U439_3p	U439_3p	5.36	0.00	Up	not DE	Down
U45_5p	hsa-miR-181c-5p	2.50	1.00	Up	Up	not DE
U46_5p	eca-miR-30c	-0.95	0.28	Down	not DE	Up
U49_3p	eca-miR-92b	-1.78	-0.93	Down	not DE	not DE
U491_3p	U491_3p	4.58	3.61	Up	not DE	not DE
U5_5p	hsa-miR-27b-5p	-1.01	-0.12	Down	not DE	Up
U62_3p	eca-miR-365	-1.39	-0.57	Down	not DE	not DE
U63_3p	eca-miR-130a	1.18	1.47	Up	Up	not DE
U65_5p	eca-let-7c	1.36	1.63	Up	Up	Up
U7_5p	eca-miR-10a	11.09	2.39	Up	Up	Down
U75_3p	eca-miR-19b	2.91	1.93	Up	Up	Down
U76_5p	eca-miR-99a	1.45	1.86	Up	Up	Up
U78_5p	eca-miR-146b-5p	2.26	-0.17	Up	not DE	Down
U83_3p	eca-miR-3958	-5.22	-0.57	Down	not DE	Up
U83_5p	oar-miR-3958-5p	-5.29	-1.44	Down	not DE	not DE
U86_5p	eca-miR-106b	2.69	1.94	Up	Up	not DE
U88_3p	eca-miR-615-3p	13.34	3.39	Up	not DE	Down
U88_5p	eca-miR-615-5p	4.75	0.00	Up	not DE	not DE
U91_5p	eca-miR-195	5.75	0.82	Up	not DE	Down
U93_3p	eca-miR-505	-1.10	-0.62	Down	not DE	not DE
U97_3p	rno-miR-196b-3p	4.76	0.00	Up	not DE	Down
U97_5p	eca-miR-196b	13.19	0.00	Up	not DE	Down
U98_5p	eca-miR-497	5.23	0.23	Up	not DE	Down

Figure 5.16. Examples of biomarkers of BPV-1 in equine fibroblasts.
Histograms 1-4 are examples of miRNAs that equine fibroblasts, in the presence of BPV-1 genomes, expressed at a low level or undetectable level. Histograms 5-8 are examples of miRNAs where presence of BPV-1 genomes was linked to increased expression. Data is based on normalised read number from HTS. Average expression from three libraries per cell lines is represented as continuous horizontal line and standard deviation is represented as dashed line

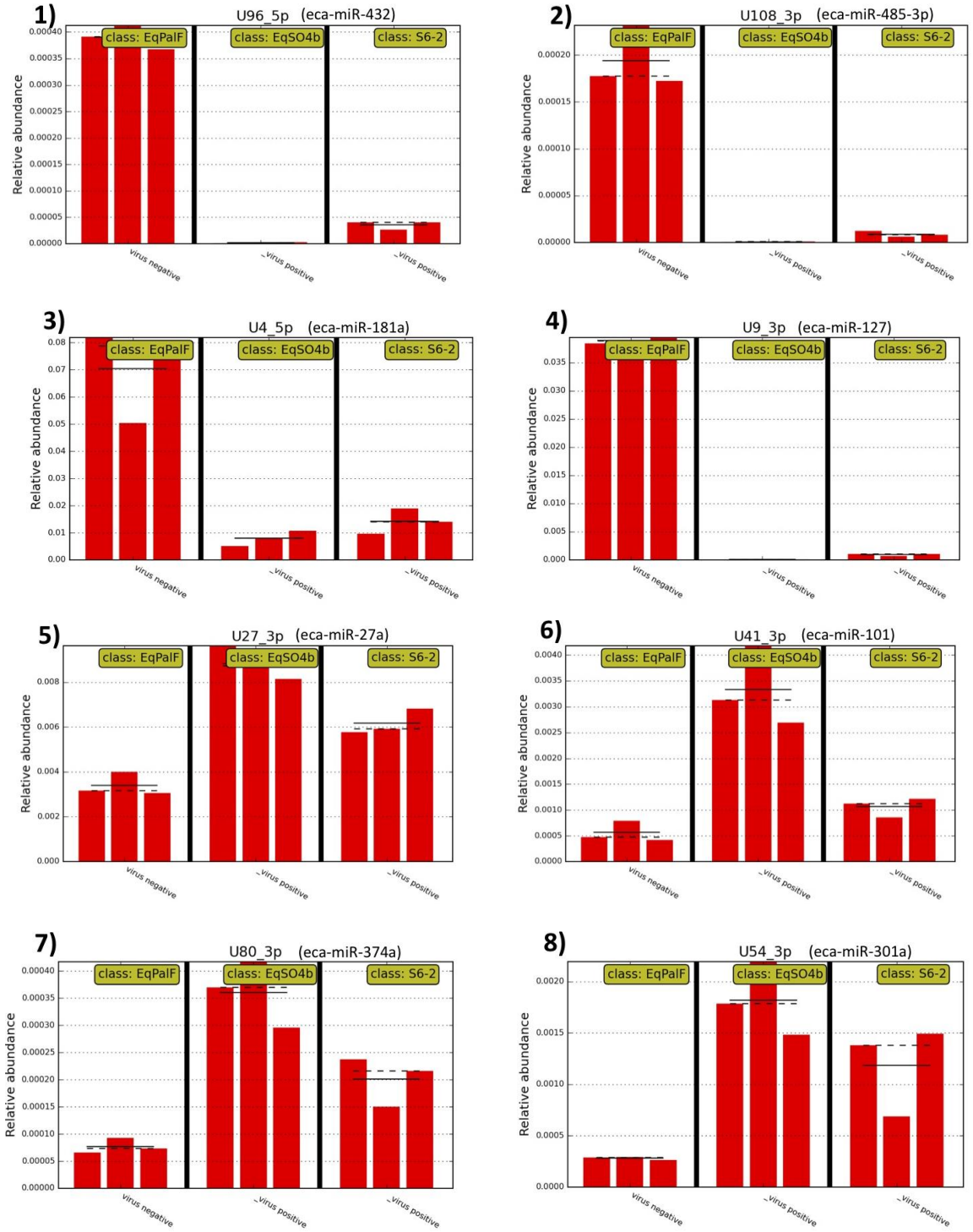
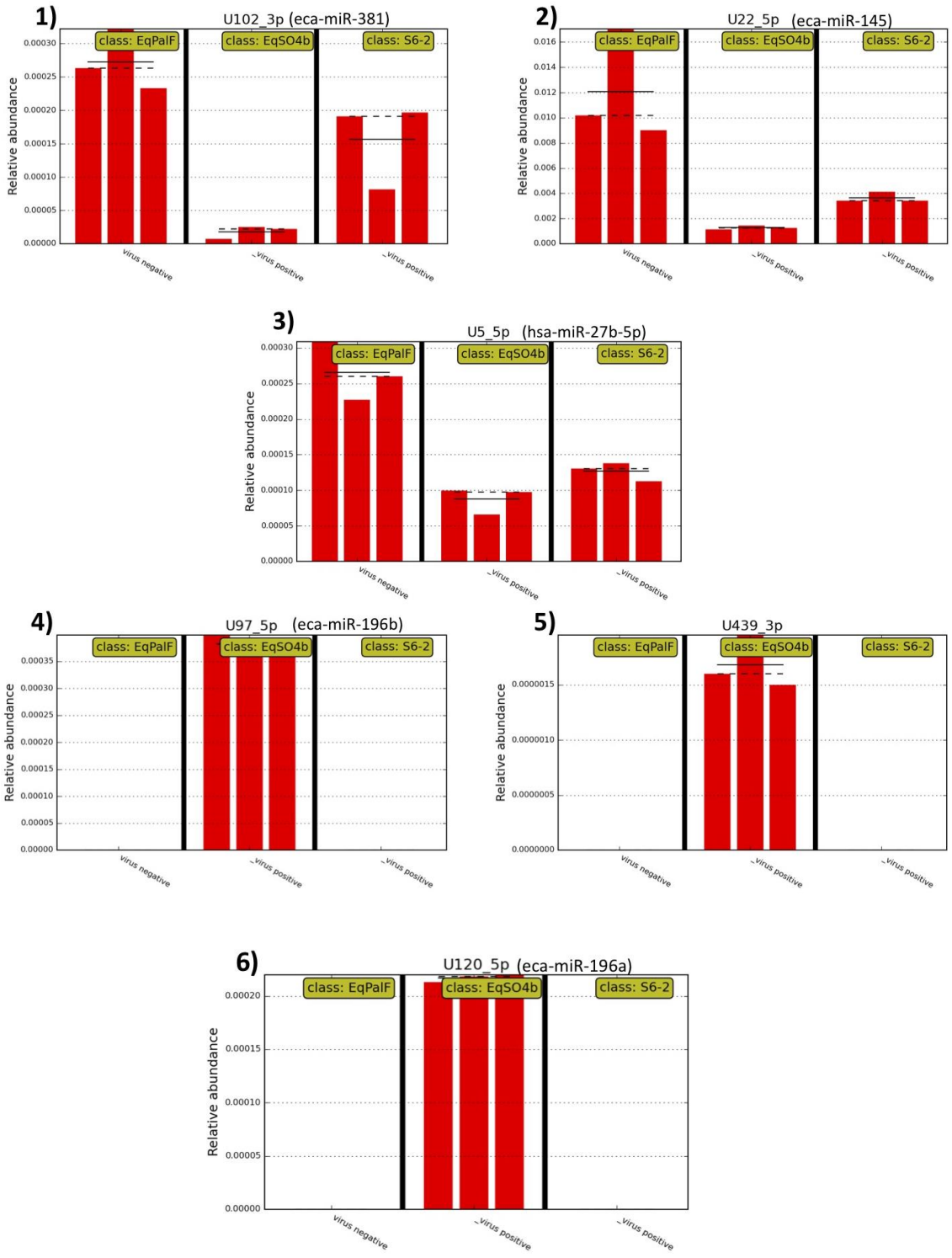


Figure 5.17. Examples of biomarkers of tumoural phenotype in equine fibroblasts. Histograms 1-3 are examples of miRNAs that equine fibroblasts, in the tumoural phenotype, expressed at a low level or undetectable level. Histograms 4-6 are examples of miRNAs of tumoural phenotype linked to increased expression. Data is based on normalised read number from HTS. Average expression from three libraries per cell line is represented as continuous horizontal line and standard deviation is represented as dashed line



5.4. Discussion

In the current chapter, the differences in microRNA profiles associated with ES were explored using HTS in three cell lines that represent an *in vitro* model of equine sarcoids, namely EqPalF cells (equine primary fibroblasts), S6-2 cells (equine primary fibroblasts *in vitro* transformed with BPV-1 genomes and expressing BPV-1 oncoprotein E5) and EqSO4b cells (equine fibroblasts from a naturally occurring sarcoid that have episomal BPV-1 genomes and express BPV-1 oncoprotein E5) (Yuan *et al.*, 2008a). Bioinformatic tools identified a group of putative novel miRNAs in equine fibroblasts (285), miRNAs linked to BPV-1 presence (201), miRNAs linked to tumoural transformation (128) and miRNAs with potential use as biomarkers for the presence of BPV-1 and/or tumoural development (106). Novel miRNAs will be submitted to miRBase (http://mirbase.org/) to confirm their identity and they will represent an important contribution to the scientific community as a resource for further studies on equine miRNAs.

Data obtained from the nine libraries representing the three cell lines was considered robust as RNA was of good quality and linear correlation of read counts between replicates was strong. Additionally, results for four out of the six miRNAs (miR-127, miR-196b, miR-409-3p and miR-410) quantified using Taqman® qRT-PCR were in agreement with the results obtained by HTS, quantifying differential expression among the three cell lines. MA and PCA plots indicated that there is a clear biological differences between the three cell lines, grouping EqSO4b and S6-2 together suggesting that the BPV-1 positive cells are more similar to each other than to EqPalF (Figures 5.7-5.8). In the pair-wise comparisons, the pattern of expression of miRNAs is clearly different between EqPalF and the two cell lines that contain BPV-1 genomes, with some miRNAs showing substantial differences in levels of expression. In some cases this marked upregulation or downregulation was due to the lack of expression or very low expression of some miRNAs in one or more of the cell lines. This was the case for U120-5p (eca-miR-196a) and U97_5p (eca-miR-196b) that were not expressed in S6-2 cells nor EqPalF cells but had over 4,000 reads in EqSO4b (average of three libraries). miRNA U154-3p (eca-miR-296) was marginally

expressed in EqSO4b and absent in S6-2 cells but had an intermediate level of 955 reads in EqPalF.

A higher number of miRNAs were found dysregulated between tumoural cells and equine primary fibroblasts (329) than between equine primary fibroblasts *in vitro* transformed with BPV-1 genomes and equine primary fibroblasts (251). This is unsurprising as, although the miRNA profile in S6-2 cells and EqSO4b is more similar to each other than to EqPalF, S6-2 cells were derived from EqPalF (Yuan *et al.*, 2008a). Interestingly, there was strong correlation between the levels of expression of dysregulated miRNAs between the two cells lines containing BPV-1 genomes (please refer to Figure 5.14). In fact, only three miRNAs showed differences between the two cell lines. The rest, 198 (99%) were commonly regulated in S6-2 and EqSO4b cells. This high degree of correlation indicates that BPV-1 is likely to play an important role in miRNA dysregulation in the *in vitro* model of ES utilised in the current study.

In recent years, multiple HTS studies performed in *in vitro* and *ex vivo* HPV-16/-18 positive specimens highlighted differences associated with HPV infection and tumour development (Witten *et al.*, 2010; Greco *et al.*, 2011; Wang *et al.*, 2014b; Honegger *et al.*, 2015) and many dysregulated miRNAs are known to be linked to the expression of HR-HPV oncoproteins (see Table 5.1). Of the miRNAs dysregulated in BPV-1 positive cells (201), when compared to primary fibroblasts, a similar number were found to be upregulated (102) and downregulated (96). The 201 miRNAs commonly dysregulated in BPV-1 positive cells compared to equine primary fibroblasts may be attributed to the presence of episomal BPV-1 genomes and BPV-1 oncoprotein expression. Of the 201, the 20 miRNAs that were significantly higher in the tumoural cell line (EqSO4b) and the 43 that were significantly lower in the tumoural cell line when compared to BPV-1 transformed cells (S6-2), may play an important role in neoplastic transformation and cellular pathology associated with tumorigenesis.

In the group of miRNAs upregulated in BPV-1 positive cells, several members of the miR-17~92 cluster were found (miR-17, miR-18a, miR-19a, miR-19b, miR-20 and miR-92a). Members of this cluster have been reported to be upregulated as a consequence of HPV c-Myc and p21 activation and some of them such as miR-19a, miR-19b and miR-20a, have validated targets that infer tumorigenic activity

in HPV-induced cancer (Sharma *et al.*, 2014; Diaz-Gonzalez *et al.*, 2015). It has been suggested that upregulation of these miRNAs in HPV-16/-18 induced cancer may be related with increased expression of c-Myc due to HR-HPV E7 oncoprotein activation (Gómez-Gómez *et al.*, 2013). BPV-1 E7 does not have a domain to activate pRB directly, which is how HPV-16/-18 E7 activates c-Myc, but it can activate p600, a member of the retinoblastoma family (DeMasi *et al.*, 2005; Corteggio *et al.*, 2011). It is possible that BPV-1 E7 may trigger similar but not identical mechanisms to HPV to upregulate these miRNAs. Upregulation of another miRNA from this cluster, eca-miR-92a in both BPV-1 positive cell lines, may be linked to the expression of BPV-1 oncoproteins since the human homologue of eca-miR-92a, hsa-miR-92a is thought to contribute to tumour development and is upregulated by HPV-16 E6 (via PTEN protein) (Yu *et al.*, 2013). Further experiments will be necessary to corroborate how BPV-1 upregulates some of the members of the miR-17~92 cluster.

Other miRNAs of interest in the upregulated group were miR-106b, miR-15a, miR-16 and miR-21. These four miRNAs have been found to be upregulated in HR-HPV induced cancer and have been linked to the expression of HPV oncoproteins E6 and E7 (Wang *et al.*, 2008; Pereira *et al.*, 2010; Li *et al.*, 2011; Lajer *et al.*, 2012; Wang *et al.*, 2014b). It is unclear whether upregulation of miR-106b in HPV occurs via E6 or E7 but it leads to an increase in cell migration (Gómez-Gómez *et al.*, 2013). miRNAs from the 15/16 cluster (miR-15a and miR-16) on the other hand, have been reported to be upregulated via c-Myc, again by HPV E7 activation. They are thought to promote angiogenesis and metastasis (Myklebust *et al.*, 2011). miR-21 is a well-known oncomiR that has been reported to be upregulated in numerous types of neoplasias and in HPV-related cervical cancer through the expression of E6 oncoprotein (Ben *et al.*, 2015; Xu *et al.*, 2015). Excess miR-21 promotes proliferation, cell migration and invasion in cervical cancer cells by increasing phosphorylation of Akt and degradation of tumour suppressor PTEN (Ben *et al.*, 2015). Further studies are needed to ascertain the roles and target genes of miR-106b, miR-15-16 cluster and miR-21 in BPV-1 tumorigenesis.

Amongst the 43 most downregulated miRNAs in BPV-1 positive cells when compared to equine primary fibroblasts, miR-181a, miR-181b and miR-204b were

of particular interest. miR-181a and miR-181b were found to be downregulated by both methods (HTS and microarray) in BPV-1 positive cells and miR-181a showed the same pattern by qRT-PCR (see section 3.3.3. Chapter 3). Importantly, low levels of miR-181a were flagged up by the algorithm in LEfSe as a suitable biomarker for BPV-1 presence. Interestingly, the human homologue hsa-miR-181a has been reported to be downregulated in HR-HPV *in vitro* studies by microarray and qRT-PCR (Wang *et al.*, 2008; Lajer *et al.*, 2011; Wald *et al.*, 2011) and there is evidence that HPV-16 inhibits the expression of this miRNA by directly preventing its transcription (Lee *et al.*, 2015). What effect reduced levels of miR-181a have on fibroblast cells will require further examination, for example with the use of miRNA inhibitors. Both miR-181a and b were found to be encoded in a region of the equine genome that has been associated with increased risk of ES (Jandova *et al.*, 2012).

miR-204b was also observed to be downregulated in BPV-1 positive cells using two techniques (HTS and microarray). The homologue hsa-miR-204b-5p was previously reported from HTS data to be downregulated in HR-HPV-induced cancer (Witten *et al.*, 2010) although the mechanisms by which this occurs remain unknown. A recent genetic study has reported deletions, amplifications and polymorphisms associated with predisposition to equine sarcoids (Staiger *et al.*, 2016) and some of these mechanisms may explain aberrant expression associated with HPV-induced cervical cancer (Sharma *et al.*, 2014). Notably, miR-204b is encoded in a region of the equine genome recently identified as being deleted in horses with ES (Bugno-Poniewierska *et al.*, 2016). miR-181a/b and miR-204 may be examples of deletions triggered by BPV-1 during tumorigenesis and may be useful to determine tumorigenic status in ES as they appear to be depleted in both BPV-1 transformed and tumoural cells. Further studies using clinical samples would help determine the usefulness of these miRNAs as biomarkers for ES and to ascertain the mechanisms by which BPV-1 triggers their downregulation. In addition, biomarkers may help identify whether a change in expression of any single miRNA can lead to ES or if dysregulation of multiple miRNAs is involved.

In total, 128 miRNAs were dysregulated in the tumoural cells and not in the BPV-1 transformed cells (S6-2). Amongst the upregulated miRNAs miR-16-3p, miR-

106a, miR-20a-3p, miR-196a-3p/5p and miR-196b-3p/5p were of interest. miR-16-3p is part of cluster miR-15/-16 and miR-20a and miR-106a are part of cluster 17~92. Other members of these two clusters were found upregulated in both BPV-1 positive cells of the ES *in vitro* model (EqSO4b and S6-2 cells), and both clusters have been linked to expression of HPV E6 and E7 oncoproteins in the literature as discussed above. Currently it is unclear what the mechanisms are behind the upregulation of these clusters in ES but it is likely that they are linked to BPV-1 oncoproteins. miR-196a and miR-196b on the other hand are two miRNAs that have been linked to tumorigenesis and invasion and reported to be upregulated in HPV-induced cancer and suggested as a biomarker for prognosis (Pereira *et al.*, 2010; Gocze *et al.*, 2013; Lu *et al.*, 2014; Vojtechova *et al.*, 2016). Higher expression of miR-196a in HPV-induced cervical cancer has been associated with expression of oncoprotein E7 via the PI3K pathway (Pereira *et al.*, 2010). Both miR-196a and b have been associated with increased levels of MMPs through activation of the NME4-JNK-TIMP1-MMPs pathway in oral cancer, and in colon cancer miR-196b has been reported to regulate apoptosis by targeting Fas receptor (CD95) (Lu *et al.*, 2014; Mo *et al.*, 2015). In the *in vitro* model used for the current project, EqPalF and S6-2 cells were depleted of miR-196a and miR-196b. In contrast, levels of miR-196b were shown to be significantly upregulated in EqSO4b cells by HTS and qRT-PCR and both miR-196a and miR-196b were identified as potential biomarkers for tumoural phenotype by the LEfSe algorithm. In ES, the PI3K pathway has been reported to be upregulated via activation of PDGFß-R by BPV-1 E5 and upregulation of MMPs is linked to expression of BPV-1 E6 and E7 (Borzacchiello *et al.*, 2009; Yuan *et al.*, 2010b; Yuan *et al.*, 2011a). It is possible that BPV-1 regulates miR-196a and b in a similar way to HPV and it is likely that the upregulation of these miRNAs contributes to the upregulation of MMPs registered in ES (Yuan *et al.*, 2010b).

Amongst the miRNAs downregulated in EqSO4b cells and not in S6-2 cells, miR-23a, miR-23b and miR-27b-5p were of interest as they form part of a cluster that has also been found to be downregulated in HPV-16/-18 induced cancer (Wang *et al.*, 2008; Greco *et al.*, 2011; Yeung *et al.*, 2011; Honegger *et al.*, 2015). HPV-16 E6 decreases the amount of tumour suppressor p53 by degrading it and, as a consequence of p53 deactivation and degradation, multiple miRNAs are downregulated such as miR-145, miR-23b, miR-218 and miR-34a (McBee *et al.*,

2011; Yeung *et al.*, 2011; Shi *et al.*, 2012; Diaz-Gonzalez *et al.*, 2015). A decrease in levels of miRNAs such as miR-23b has an effect on cell migration and metastasis (Sharma *et al.*, 2014). Although BPV-1 downregulates p53 in a different way to HPV, it is likely that mislocation of p53 by BPV-1 E6 will have an effect on miRNAs that are regulated by p53 such as miR-23b and therefore contribute to the tumoral phenotype seen in EqSO4b cells.

Functional analysis of the differentially expressed miRNAs in the two BPV-1 positive cell lines identified almost identical networks, diseases and biofunctions for both cell lines. Interestingly these were related to cancer, reproductive system (HPV), connective tissue and cellular mechanisms that participate in tumorigenesis (such as cell proliferation and cell cycle). The findings of the analysis fit with the disease model of ES, although it needs to be said that multiple miRNAs were excluded as no homologous annotation was available in miRBase 20.0 or they were novel miRNAs. Some of the miRNAs included for IPA were 'transformed' to their related miRNA in human or mice and the cut-off point for mismatches was set at 2 nucleotides (either in length or composition). Although some miRNAs could not be included in the analysis and there were dissimilarities between the miRNAs found in equine fibroblasts and the miRNAs used as the input for IPA analysis, interesting miRNAs were highlighted by the software due to their involvement in multiple networks and/or because multiple dysregulated miRNAs were members of the same network. Several miRNAs were identified by the algorithm of LEfSe as suitable biomarkers for the presence of BPV-1 and biomarkers for tumoural phenotype and are presented in section 5.3.9. LEfSe is a free online tool with the advantage over other tools that it does not require the mature sequence of a miRNA or the annotation from miRBAse as the calculations are based purely on the levels of expression of the miRNAs. The software is not specific for miRNAs and it has been used to suggest potential biomarkers in periodontitis, inflammatory bowel disease and cervical cancer (Davenport *et al.*, 2014; Szafranski *et al.*, 2015; Dareng *et al.*, 2016). Although the list produced by the software identified 59 miRNAs for BPV-1 presence and 47 miRNAs for tumoural phenotype, several miRNAs showed the same pattern of expression (displayed in Figure 5.16 and Figure 5.17). Tumorigenesis is associated with changes in expression of multiple miRNAs rather than one unique miRNA (Jansson & Lund, 2012; Liu, 2013). Further analysis using clinical tumour

samples will be required to assess the utility of the suggested biomarkers in ES and perhaps to refine the list based on *in vivo* expression levels.

Comparisons to microarray results and to data from previous studies of miRNAs in HPV-induce cancer were challenging. Firstly, numerous miRNAs found in BPV-1 positive equine fibroblasts had no annotation in miRBase (novel miRNA candidates). Secondly, differences were present between the nucleotide composition and/or length of miRNAs obtained with HTS and the probes used in the microarray. Lastly, there is no centralised database for miRNAs dysregulated in papillomavirus related diseases. The closest found was HPVbase although this database did not include many miRNAs. Four review studies together with the results from 13 publications on HR-HPV related cancer dysregulated miRNAs from *in vitro* and *ex vivo* samples were utilised to construct a database to compare the data from HTS obtained from the ES *in vitro* model from the current study (Wang *et al.*, 2008; Pereira *et al.*, 2010; Witten *et al.*, 2010; Greco *et al.*, 2011; Li *et al.*, 2011; McBee *et al.*, 2011; Wald *et al.*, 2011; Yeung *et al.*, 2011; Zheng & Wang, 2011; Lajer *et al.*, 2012; Gocze *et al.*, 2013; Gómez-Gómez *et al.*, 2013; Sharma *et al.*, 2014; Wang *et al.*, 2014b; Ben *et al.*, 2015; Honegger *et al.*, 2015; Wang *et al.*, 2016).

Discrepancies between the results obtained by microarray from S6-2 *vs.* EqPalF cells and HTS may be attributable to the differences in sensitivity between microarray and HTS as HTS can detect very low levels of expression. The probes utilised for microarray were from an external company using equine predicted sequences and human sequences and therefore mismatches between these and the miRNAs in equine fibroblasts may have reduced the hybridisation signal detectable. This type of discrepancy has been recognised in previous studies where HPV+ cells/tissues also showed disparity between techniques used to assess miRNA expression (reviewed by Sharma *et al.* (2014)). Additionally, although the EqPalF RNA sample for both microarray and HTS was obtained from the same culture and RNA extraction (technical replicates), the RNA obtained from S6-2 cells for the microarray and for the HTS was from four different cell cultures (biological replicates). It is well known that in papillomavirus infection, miRNA profile changes with time and miRNAs that may be expressed in low amounts in early stages may increase with time or vice versa (Greco *et al.*, 2011;

Li *et al.*, 2011). Examples of this effect are miR-106a, miR-92a and miR-375 (Li *et al.*, 2011). Despite this, correlation between libraries of each cell line was excellent which accounts the robustness of the data present in this chapter.

Despite difficulties in the comparison between miRNAs found in BPV-1 positive cells and HR-HPV studies, and the inherent differences in the tumoural transformation induced by each virus (Moody & Laimins, 2010; Corteggio *et al.*, 2011; Venuti *et al.*, 2011; Gil da Costa & Medeiros, 2014; Munday, 2014), multiple miRNAs were found dysregulated in BPV-1 positive cells in the current study that were also dysregulated in the database elaborated with information from HR-HPV induced cancer studies. This was the case for miR-23b, miR-181a, miR-191a, miR-145, miR-204b and miR-409-3p; miR-10a, miR-15a, miR-16 and miR-21 that were found downregulated and upregulated respectively in both BPV-1 positive equine fibroblasts and HPV-16/-18 *in vitro* and *ex vivo* studies (Wang *et al.*, 2008; Pereira *et al.*, 2010; Witten *et al.*, 2010; Lajer *et al.*, 2011; Li *et al.*, 2011; McBee *et al.*, 2011; Yeung *et al.*, 2011; Lajer *et al.*, 2012; Wang *et al.*, 2014b). In other cases such as miR-15b, miR-92a, miR-99a and miR-132, opposite direction of dysregulation was noticed between BPV-1 positive cells and HPV-16/-18 cancer cell lines and cervical cancer samples (Wang *et al.*, 2008; Pereira *et al.*, 2010; Witten *et al.*, 2010; Li *et al.*, 2011; Lajer *et al.*, 2012). Regardless of differences in the direction of expression, the comparison to HR-HPV was useful and suggests that there may be miRNAs regulated in a similar manner between BPV-1 in ES and HR-HPV.

Regarding data analysis, no unique model or pipeline for miRNA description and miRNA discovery was found in the literature. Therefore several assumptions were made during data mining. In the first place, during pre-processing any sequence found five or more times in the equine genome was excluded from the analysis. In removing ubiquitous sequences, potential paralogous may have been lost. Additionally, for the annotation methods (BLAST and Edit distance, described in section 2.5.2, Chapter 2), taking two as a cut-off point to define different miRNAs may have introduced errors in annotation. This cut-off point was applied after consulting with external companies and assessing the data with different cut-off points. It was observed that the majority of the miRNAs when compared to miRBase 20.0 were either identical or had up to two or three

differences. Where four or more differences in nucleotides (in length or in composition) sequences were considered to be unrelated miRNAs to the sequence being interrogated. Based on these conclusions, two differences in length or composition were set as the limit. This was deemed important since functional analysis can only be performed using annotated miRNAs once sequence data have been accepted for publication. Following acceptance of the 752 sequences of miRNAs candidates found in equine fibroblasts by miRBase, additional functional analysis will be possible as the tools for functional analysis rely on miRNA annotations from the miRBase, either in the form of eca-miR-xx-5p/3p or MIMA-XXXXX.

Another limitation found during post-processing was selecting the criteria of 'presence of miRNAs in at least two of the replicates of each cell line' to be considered a 'real' miRNA, irrespectively of the read count. Some publications have used a minimum number of reads but since this study was the first to characterise miRNAs expressed in equine fibroblasts, utilising a minimum number of reads would have excluded miRNAs expressed at low levels. Presence in multiple replicates from the three cell lines rather than a minimum number of reads was thought more appropriate in this case, although it would have left out miRNAs that were only present in one replicate of each cell line. When the distribution of the reads across replicates and cell lines was initially studied, all but one of the miRNAs left out by the 'at least present in two replicates' filter had a read count lower than 1,000, thus supporting the appropriateness methodology.

In summary, this chapter describes the differential expression of miRNAs in an ES *in vitro* sarcoid model with the use of HTS. Results were validated with qRT-PCR and compared to previous microarray data and previous studies on miRNAs in HR-HPV-induced cancer. In total, 285 novel miRNAs were identified and 201 and 128 miRNAs were found to be linked to BPV-1 presence and tumoural development respectively. Further studies assessing potential cellular targets and the role of BPV-1 oncoproteins in the dysregulation of these miRNAs will be warranted to understand the roles of these miRNAs in the tumorigenesis process.

Chapter 6. General discussion

6.1. Introduction

In this final discussion, I will review how the aims of this PhD project were met, followed by a summary of the major discoveries described in each chapter. I will compare the results generated using different techniques and the limitations of each approach. Lastly, reflection on how findings from this project have contributed to our knowledge of equine miRNAs and equine sarcoids and what objectives future work should address, are presented.

6.2. How the aims were met, comparative analysis of the results and limitations of the approaches used

This project aimed to improve knowledge of equine miRNAs and provide a better understanding of the miRNAs dysregulated in equine sarcoids using an *in vitro* model with BPV-1 transformed equine fibroblasts.

The miRNA profile in normal equine fibroblasts (EqPalF cells) was assessed using two different techniques: microarray and high throughput sequencing (HTS). The microarray included probes that would potentially hybridise to predicted equine mature miRNAs, human mature miRNAs, and a small number of non-papillomavirus miRNAs, synthetic miRNAs and small nucleolar RNA sequences. This technique was fast and required a very small amount of RNA from the cells therefore proved to be useful for an initial identification of miRNAs in equine fibroblasts. The drawbacks of the microarray technique were that multiple miRNAs with low levels of expression were not detected and the hybridisation may have not been as efficient as expected due to sequence differences between the expressed and predicted miRNAs on the array. Nevertheless, a total of 492 probes out of 2164 on the array showed hybridization. One quarter of the probes (124) corresponded to predicted equine miRNA sequences deposited in miRBase 21, and detection of these by microarray provided experimental evidence that these predicted sequences are indeed expressed in horse fibroblast cells.

Extension of this initial analysis using HTS provided a more detailed description of all miRNAs expressed in these cells. As expected, HTS had a higher dynamic range than microarray which allowed detection of miRNAs expressed at low levels. In addition, as HTS does not require previous knowledge of the sequence of the mature miRNAs, discovery of novel miRNAs was possible. The drawbacks of the HTS technique were that it required a higher amount of much higher quality RNA and involved intensive data analysis. Nevertheless, this approach identified 593 miRNAs expressed in equine fibroblasts, of which 335 were never described in the horse before. However, although bioinformatic analysis of the HTS results was carried out in detail, from this analysis alone it cannot be concluded that these are true, novel miRNAs; further clarification of this and naming of new miRNAs will be provided by the miRBase curators who will test if the read pattern provides evidence for processing by Drosha and Dicer with criteria such as mismatches on the sequence between mature forms and hairpin precursor, abundance of mature forms from both arms of the hairpin, folding energy and mismatches with the predicted sequences (Kozomara & Griffiths-Jones, 2014). However, HTS did provide data on a number of interesting aspects. Firstly, equine miRNA length and GC content were found to be highly similar to what was previously predicted in the horse and suggested in previous studies on equine miRNAs (Zhou *et al.*, 2009; Desjardin *et al.*, 2014; Kim *et al.*, 2014). Secondly, all equine chromosomes (ECA), with the exception of one (ECA 31), encoded miRNAs and interestingly ECA 24 encoded a high number and high density of miRNAs. miRNAs from equine fibroblast cells were commonly expressed in clusters of between 2 and 13 miRNAs and 6.75% of the miRNAs expressed mapped to multiple regions in the genome, indicating gene duplication. The range of expression of those miRNAs was variable and the top 50 most abundant miRNAs found in clusters had reads above 1000 (average reads for three replicates of EqPalF). Finally, 258 predicted miRNAs and 335 novel miRNAs were verified in equine primary fibroblasts. Of these novel miRNAs, 162 sequences were potentially new equine miRNAs and 173 sequences were potentially novel miRNAs, not previously described in any species.

To address the second aim of the project, the study of differential expression of miRNAs in a well-established *in vitro* model of ES, three state-of-the-art technologies were utilised: microarray, HTS and real-time quantitative reverse-

transcription polymerase chain reaction (qRT-PCR). The *in vitro* model, developed at the University of Glasgow, included an *in vitro* transformed cell line with BPV-1 episomal genomes (S6-2; equine primary fibroblast containing BPV-1 genomes and expressing BPV-1 oncoproteins), a tumoural cell line (EqS04b; equine fibroblasts obtained from naturally occurring sarcoids, containing BPV-1 genomes and expressing BPV-1 oncoproteins) (Yuan *et al.*, 2008a) and a control cell line (EqPalF; equine primary fibroblasts). Results from the microarray showed that 206 miRNAs were differentially expressed between the control and *in vitro* transformed cell lines (statistical significance at $p<0.05$). Interestingly, more than two thirds of the miRNAs interrogated on the microarray (70%) were downregulated as a consequence of the presence of BPV-1 genome. There are two possible explanations for this result. Firstly, a true biological reason may be responsible for this general downregulation of miRNAs, such as the presence of BPV-1 episomal genomes. Secondly, possible bias may have been introduced by the set of probes that are included on the microarray, the list of which is dictated by the external company. When the results from the three cell lines of the model were examined using HTS a different profile of miRNAs was seen. The distribution of the dysregulated miRNAs was relatively even: approximately 50% of the miRNAs were downregulated when comparing BPV-1 positive cells (either *in vitro* transformed or tumoural cells) to control cells. Therefore, the finding from the microarray data that the presence of BPV-1 genomes led to general downregulation in the expression of miRNAs in equine fibroblasts was not supported by the HTS data. In the light of the HTS results it is hypothesised that this trend was due to technical reasons such as the bias in the set probes. The highly similar miRNA expression profiles in both BPV-1 positive cell lines (S6-2 and EqSO4b) strengthens this hypothesis because similar results were found when comparing these to the control cell line. Furthermore, analysis of the HTS data showed a number of interesting discoveries. Firstly, different biological replicates (different group of cultured cells and different RNA preparations) showed excellent correlation in the miRNAs expressed and in level of expression. This was an expected finding as there should be no major differences in the culture procedure and RNA extraction. However it added robustness to the data set. Secondly, by comparing the miRNA profiles, higher similarities were found between the *in vitro* transformed cells and the tumoural cells than between the *in vitro* transformed cells and the control cells. This was

perhaps an unexpected finding since transformed cells are descendants of the control cells. Clearly, the presence of BPV-1 genomes seems to be sufficient to alter the miRNA expression profile of in vitro transformed cells and make these more similar to tumoural cells, as shown in Figure 5.8. (Chapter 5). Additionally, as there was a strong correlation between the expression and types of miRNAs dysregulated in both BPV-1 positive cells relative to control cells, the dysregulation of those miRNAs could be directly attributed to the presence of BPV-1 virus and to the expression of BPV-1 oncoproteins. Thirdly, a subset of 201 miRNAs was found to be dysregulated due to the presence of the BPV-1 genome and amongst these, several miRNAs had been previously linked to the tumoural transformation driven by high-risk human papillomaviruses (HR-HPVs) oncoproteins. Three miRNAs within this group, U4_5p (miR-181a), U44_5p (miR-181b) and U72_5p (miR-204b) were also encoded in areas of the equine genome previously reported to be linked to ES (Jandova *et al.*, 2012; Bugno-Poniewierska *et al.*, 2016). miR-181a and miR-181b were found to be downregulated by both HTS and microarray in BPV-1 positive cells and miR-181a showed the same pattern by qRT-PCR. Importantly, low levels of miR-181a were flagged up by the algorithm in LEfSe as a suitable biomarker for BPV-1 presence. Interestingly, the human homologue hsa-miR-181a has been reported to be downregulated in HR-HPV *in vitro* studies by microarray and qRT-PCR (Wang *et al.*, 2008; Lajer *et al.*, 2011; Wald *et al.*, 2011) and there is evidence that HPV-16 inhibits the expression of this miRNA by directly preventing its transcription (Lee *et al.*, 2015). Nevertheless, the roles of these three miRNAs in ES warrants further investigation. Studies assessing tumoural transformation parameters in miR-181a/b and miR-204b knocked-down cell lines and quantification of those miRNAs in BPV-1 E5, E6 and E7 transformed cells would help to clarify the roles of those miRNAs in the *in vivo* BPV-1 tumoural transformation. Quantification of those miRNAs in the six different clinical types of equine sarcoids using qRT-PCR may help to understand their role in the disease.

Another interesting finding was a subset of 128 miRNAs that were found to be linked to the tumoural phenotype. These were not dysregulated in BPV-1 transformed cells but were significantly dysregulated in tumoural cells when compared to control cells. Amongst these miRNAs, several were reported to be linked to HR-HPV oncogenic transformation. In particular, miR-196a and b were

reported to be associated with increased levels of MMPs through activation of NME4-JNK-TIMP1-MMPs pathway in oral cancer, and in colon cancer miR-196b has been reported to regulate apoptosis by targeting Fas receptor (CD95) (Lu *et al.*, 2014; Mo *et al.*, 2015). Higher expression of miR-196a in HPV-induced cervical cancer has been associated with expression of oncoprotein E7 via PI3K pathway (Pereira *et al.*, 2010) and it has been suggested as a biomarker for prognosis in this disease (Pereira *et al.*, 2010; Gocze *et al.*, 2013; Lu *et al.*, 2014; Vojtechova *et al.*, 2016). Finally, the use of bioinformatic tools for pathway/network analysis and biomarker prediction proved to be informative in the *in vitro* ES model, as the major networks identified fit the disease model for ES (cancer, reproductive tract disease linked to HPV and tumorigenesis) and several miRNAs considered relevant, due to their association with HR-HPV cancer or other types of cancer, were predicted as suitable biomarkers. Predicted biomarkers for BPV-1 presence were miR-92a (member of the miR-17~92 cluster), miRNAs from the 15/16 cluster and miR-21. miR-92a in HR-HPV is thought to contribute to tumour development and is upregulated by HPV-16 E6 (via PTEN protein) (Yu *et al.*, 2013), miRNAs from the 15/16 cluster have been reported to be upregulated via c-Myc, again by HPV E7 activation and miR-21 is a well-known oncomiR that has been reported to be upregulated in numerous types of neoplasias and in HPV-related cervical cancer through the expression of E6 oncoprotein (Ben *et al.*, 2015; Xu *et al.*, 2015). Following submission of the subset of miRNAs identified in equine fibroblasts to the official miRNA database, a more comprehensive functional analysis will be warranted; currently the tools for functional analysis list only a small number of miRNAs found in equine fibroblasts in the current study. Results from these predictive tools will need to be validated with *ex vivo* specimens to assess the clinical usefulness of these miRNAs as potential biomarkers.

6.3. Contribution of these findings to the current knowledge of ES and equine miRNAs.

Bioinformatic analysis of the data obtained by HTS and validation using qRT-PCR showed that 280 of the predicted miRNAs listed in miRBase for the horse are indeed expressed in equine fibroblasts (including BPV-1 positive equine fibroblasts). Of these, 42 miRNAs were exclusively expressed in the two cell lines

with BPV-1 genomes and expressing BPV-1 oncoproteins. Furthermore, 187 miRNAs (162 from EqPalF cells, and 25 from BPV-1 positive cells) that have not been reported in the horse but have orthologous in other species (novel in the horse); and 285 miRNAs (173 found in EqPalF and 112 found in BPV-1 positive cells) that have not been described in any other species (novel miRNAs) were found in equine fibroblasts. These results represent the first experimental verification that the 280 predicted miRNAs are indeed expressed in the horse and demonstrate expression of 285 additional novel miRNAs in horse fibroblast cells, which are also conserved in other species. These findings, when deposited in miRBase, will be a substantial contribution to the scientific community and will constitute a starting point for further studies in equine miRNAs.

With the use of three different techniques (microarray, HTS and qRT-PCR) 201 miRNAs were found to be significantly dysregulated in BPV-1 positive cells, attributable to the presence of BPV-1 genomes and BPV-1 oncoprotein expression. A subset of 59 miRNAs were predicted by linear discrimination analysis algorithm as suitable candidates for biomarkers of presence of BPV-1. Furthermore, 128 miRNAs were linked to tumoural transformation and of those the aforementioned algorithm predicted 47 to be suitable candidates for biomarkers of tumoural transformation. Interestingly, within both groups there were miRNAs linked to BPV-1 presence and miRNAs linked to tumoural transformation which are also reported to be linked to HR-HPV oncogenesis. These two subsets of miRNAs represent the proof that there is dysregulation of miRNAs in ES. E*x vivo* analysis of miRNAs expressed in clinical ES samples, together with histopathological characterization, is ongoing and the data will be important in testing the hypothesis that these miRNAs may be useful as diagnostic and prognostic tools. Further studies corroborating the *in vivo* results in *ex vivo* specimens with the use of HTS and qRT-PCR validation would help to understand the miRNA changes attributed to BPV-1 in the live disease. Future *in vitro* studies adding miRNA mimics or inhibitors of the miRNAs summarised above to test their effects on gene expression and phenotype will be required to fully understand the consequences of miRNA dysregulation in ES oncogenesis and will be an initial step to therapeutic options. Further studies assessing the miRNA expression patterns in different clinical types would be required to potentially developing accurate tools to aid with diagnosis and prognosis of ES.

6.4. Conclusions

Understanding of the oncogenic mechanisms of BPV-1/-2 in equine sarcoids remains an important goal to create future diagnostic, prognostic and therapeutic tools. The combination of several molecular biology and bioinformatic technologies has proven to be an efficient way to verify 280 previously predicted miRNAs, to discover 285 novel miRNAs in the horse and to identify 201 dysregulated miRNAs as a consequence of BPV-1 presence and 128 dysregulated miRNAs associated to tumoural transformation in ES. These results will be an addition to the scientific community when deposited in miRBase and have contributed to a better understanding of miRNAs in the horse and BPV-1 oncogenesis in ES.

References

Altamura, G., Corteggio, A., Nasir, L., Yuan, Z. Q., Roperto, F., & Borzacchiello, G. (2013). Analysis of activated platelet-derived growth factor beta receptor and Ras-MAP kinase pathway in equine sarcoid fibroblasts. *Biomed Res Int, 2013*, 283985.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol, 215*(3), 403-410.

Altuvia, Y., Landgraf, P., Lithwick, G., Elefant, N., Pfeffer, S., Aravin, A., Brownstein, M. J., Tuschl, T., & Margalit, H. (2005). Clustering and conservation patterns of human microRNAs. *Nucleic Acids Research, 33*(8), 2697-2706.

Ambros, V. (2004). The functions of animal microRNAs. *Nature, 431*(7006), 350-355.

Amtmann, E., Muller, H., & Sauer, G. (1980). Equine connective tissue tumors contain unintegrated bovine papilloma virus DNA. *J Virol, 35*(3), 962-964.

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology, 11*(10), 1.

Angelos, J., Oppenheim, Y., Rebhun, W., Mohammed, H., & Antczak, D. F. (1988). Evaluation of breed as a risk factor for sarcoid and uveitis in horses. *Anim Genet, 19*(4), 417-425.

Angelos, J. A., Marti, E., Lazary, S., & Carmichael, L. E. (1991). Characterization of BPV-like DNA in equine sarcoids. *Arch Virol, 119*(1-2), 95-109.

Ansari, H. A., Hediger, R., Fries, R., & Stranzinger, G. (1988). Chromosomal localization of the major histocompatibility complex of the horse (ELA) by in situ hybridization. *Immunogenetics, 28*(5), 362-364.

Araibi, E. H., Marchetti, B., Ashrafi, G. H., & Campo, M. S. (2004). Downregulation of major histocompatibility complex class I in bovine papillomas. *J Gen Virol, 85*(Pt 10), 2809-2814.

Ashrafi, G. H., Tsirimonaki, E., Marchetti, B., O'Brien, P. M., Sibbet, G. J., Andrew, L., & Campo, M. S. (2002). Down-regulation of MHC class I by bovine papillomavirus E5 oncoproteins. *Oncogene, 21*(2), 248-259.

Ashrafi, G. H., Piuko, K., Burden, F., Yuan, Z., Gault, E. A., Muller, M., Trawford, A., Reid, S. W., Nasir, L., & Campo, M. S. (2008). Vaccination of sarcoid-bearing donkeys with chimeric virus-like particles of bovine papillomavirus type 1. *J Gen Virol, 89*(Pt 1), 148-157.

Auburger, G., Klinkenberg, M., Drost, J., Marcus, K., Morales-Gordo, B., Kunz, W. S., Brandt, U., Broccoli, V., Reichmann, H., Gispert, S., & Jendrach, M. (2012). Primary Skin Fibroblasts as a Model of Parkinson's Disease. *Molecular Neurobiology, 46*(1), 20-27.

Avci, C. B., Susluer, S. Y., Caglar, H. O., Balci, T., Aygunes, D., Dodurga, Y., & Gunduz, C. (2015). Genistein-induced mir-23b expression inhibits the growth of breast cancer cells. *Contemp Oncol (Pozn), 19*(1), 32-35.

Ayele, G., Feseha, G., Bojia, E., Getachew, M., Alemayehu, F., Tesfaye, M., & Farrow, M. (2007). Sarcoids: clinical epidemiology, principal effects and treatment responses. 5th International Colloquium Working Equines. 90-98.

Baek, S., Cho, K. J., Ju, H. L., Moon, H., Choi, S. H., Chung, S. I., Park, J. Y., Choi, K. H., Kim, D. Y., & Ahn, S. H. (2015). Analysis of miRNA expression patterns in human and mouse hepatocellular carcinoma cells. *Hepatology Research, 45*(13), 1331-1340.

Banno, K., Iida, M., Yanokura, M., Kisu, I., Iwata, T., Tominaga, E., Tanaka, K., & Aoki, D. (2014). MicroRNA in cervical cancer: OncomiRs and tumor suppressor miRs in diagnosis and treatment. *The Scientific World Journal, 2014*.

Barrey, E., Bonnamy, B., Barrey, E. J., Mata, X., Chaffaux, S., & Guerin, G. (2010). Muscular microRNA expressions in healthy and myopathic horses suffering from polysaccharide storage myopathy or recurrent exertional rhabdomyolysis. *Equine Vet J Suppl*(38), 303-310.

Baskerville, S., & Bartel, D. P. (2005). Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA, 11*(3), 241-247.

Baxter, M. K., McPhillips, M. G., Ozato, K., & McBride, A. A. (2005). The mitotic chromosome binding activity of the papillomavirus E2 protein correlates with interaction with the cellular chromosomal protein, Brd4. *Journal of Virology, 79*(8), 4806-4818.

Ben, W., Yang, Y., Yuan, J., Sun, J., Huang, M., Zhang, D., & Zheng, J. (2015). Human papillomavirus 16 E6 modulates the expression of host microRNAs in cervical cancer. *Taiwan J Obstet Gynecol, 54*(4), 364-370.

Benjamini, Y., & 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.

Berezikov, E. (2011). Evolution of microRNA diversity and regulation in animals. *Nature reviews genetics, 12*(12), 846-860.

Bergman, P., Ustav, M., Sedman, J., Moreno-Lopez, J., Vennstrom, B., & Pettersson, U. (1988). The E5 gene of bovine papillomavirus type 1 is sufficient for complete oncogenic transformation of mouse fibroblasts. *Oncogene, 2*(5), 453-459.

Bernard, H. U. (2006). Phylogeny and Taxonomy of Papillomavirus. In M. S. Campo (Ed.), *Papillomavirus research: From natural history to vaccine and beyond* (pp. 11). U.K.: Caister Academy.

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

Bhardwaj, A., Singh, S., & Singh, A. P. (2010). MicroRNA-based Cancer Therapeutics: Big Hope from Small RNAs. *Mol Cell Pharmacol, 2*(5), 213-219.

Bloch, N., Breen, M., & Spradbrow, P. B. (1994). Genomic sequences of bovine papillomaviruses in formalin-fixed sarcoids from Australian horses revealed by polymerase chain reaction. *Vet Microbiol, 41*(1-2), 163-172.

Bocaneti, F., Altamura, G., Corteggio, A., Velescu, E., Roperto, F., & Borzacchiello, G. (2016). Bovine Papillomavirus: New Insights into an Old Disease. *Transbound Emerg Dis, 63*(1), 14-23.

Bogaert, L., Martens, A., De Baere, C., & Gasthuys, F. (2005). Detection of bovine papillomavirus DNA on the normal skin and in the habitual surroundings of horses with and without equine sarcoids. *Res Vet Sci, 79*(3), 253-258.

Bogaert, L., Van Poucke, M., De Baere, C., Dewulf, J., Peelman, L., Ducatelle, R., Gasthuys, F., & Martens, A. (2007). Bovine papillomavirus load and mRNA expression, cell proliferation and p53 expression in four clinical types of equine sarcoid. *Journal of general virology, 88*(8), 2155-2161.

Bogaert, L., Martens, A., Van Poucke, M., Ducatelle, R., De Cock, H., Dewulf, J., De Baere, C., Peelman, L., & Gasthuys, F. (2008). High prevalence of

bovine papillomaviral DNA in the normal skin of equine sarcoid-affected and healthy horses. *Vet Microbiol, 129*(1-2), 58-68.
Bogaert, L., Martens, A., Kast, W. M., Van Marck, E., & De Cock, H. (2010). Bovine papillomavirus DNA can be detected in keratinocytes of equine sarcoid tumors. *Veterinary microbiology, 146*(3), 269-275.
Bohl, J., Hull, B., & Vande Pol, S. B. (2001). Cooperative transformation and coexpression of bovine papillomavirus type 1 E5 and E7 proteins. *J Virol, 75*(1), 513-521.
Borzacchiello, G., Russo, V., Gentile, F., Roperto, F., Venuti, A., Nitsch, L., Campo, M., & Roperto, S. (2006). Bovine papillomavirus E5 oncoprotein binds to the activated form of the platelet-derived growth factor ß receptor in naturally occurring bovine urinary bladder tumours. *Oncogene, 25*(8), 1251-1260.
Borzacchiello, G., & Roperto, F. (2008). Bovine papillomaviruses, papillomas and cancer in cattle. *Veterinary research, 39*(5), 1.
Borzacchiello, G., Mogavero, S., De Vita, G., Roperto, S., Della Salda, L., & Roperto, F. (2009). Activated platelet-derived growth factor beta receptor expression, PI3K-AKT pathway molecular analysis, and transforming signals in equine sarcoids. *Vet Pathol, 46*(4), 589-597.
Borzachiello, G., Russon, V., DellaSalda, L., Roperto, S., & Roperto, F. (2008). Expression of platelet-derived growth factor-b receptor and bovine papillomavirus E5 and E7 oncoproteins in equine sarcoid. *Journal of Comparative Pathology, 139*, 231-237.
Braaten, K. P., & Laufer, M. R. (2008). Human Papillomavirus (HPV), HPV-Related Disease, and the HPV Vaccine. *Reviews in Obstetrics and Gynecology, 1*(1), 2-10.
Brandt, S., Haralambus, R., Schoster, A., Kirnbauer, R., & Stanek, C. (2008a). Peripheral blood mononuclear cells represent a reservoir of bovine papillomavirus DNA in sarcoid-affected equines. *J Gen Virol, 89*(Pt 6), 1390-1395.
Brandt, S., Haralambus, R., Shafti-Keramat, S., Steinborn, R., Stanek, C., & Kirnbauer, R. (2008b). A subset of equine sarcoids harbours BPV-1 DNA in a complex with L1 major capsid protein. *Virology, 375*(2), 433-441.
Brandt, S., Schoster, A., Tober, R., Kainzbauer, C., Burgstaller, J. P., Haralambus, R., Steinborn, R., Hinterhofer, C., & Stanek, C. (2011a). Consistent detection of bovine papillomavirus in lesions, intact skin and peripheral blood mononuclear cells of horses affected by hoof canker. *Equine Vet J, 43*(2), 202-209.
Brandt, S., Tober, R., Corteggio, A., Burger, S., Sabitzer, S., Walter, I., Kainzbauer, C., Steinborn, R., Nasir, L., & Borzacchiello, G. (2011b). BPV-1 infection is not confined to the dermis but also involves the epidermis of equine sarcoids. *Vet Microbiol, 150*(1-2), 35-40.
Brimer, N., Lyons, C., Wallberg, A. E., & Vande Pol, S. B. (2012). Cutaneous Papillomavirus E6 oncoproteins associate with MAML1 to repress transactivation and NOTCH signaling. *Oncogene, 31*(43), 4639-4646.
Brimer, N., Wade, R., & Vande Pol, S. (2014). Interactions between E6, FAK, and GIT1 at paxillin LD4 are necessary for transformation by bovine papillomavirus 1 E6. *J Virol, 88*(17), 9927-9933.
Brostrom, H., Fahlbrink, E., Dubath, M. L., & Lazary, S. (1988). Association between equine leucocyte antigens (ELA) and equine sarcoid tumors in the population of Swedish halfbreds and some of their families. *Vet Immunol Immunopathol, 19*(3-4), 215-223.

Brostrom, H. (1995). Equine sarcoids. A clinical and epidemiological study in relation to equine leucocyte antigens (ELA). *Acta Vet Scand, 36*(2), 223-236.

Buechli, M. E., Lamarre, J., & Koch, T. G. (2013). MicroRNA-140 expression during chondrogenic differentiation of equine cord blood-derived mesenchymal stromal cells. *Stem Cells Dev, 22*(8), 1288-1296.

Buggele, W. A., & Horvath, C. M. (2013). MicroRNA Profiling of Sendai Virus-Infected A549 Cells Identifies miR-203 as an Interferon-Inducible Regulator of IFIT1/ISG56. *J Virol, 87*(16), 9260-9270.

Bugno-Poniewierska, M., Staroń, B., Potocki, L., Gurgul, A., & Wnuk, M. (2016). Identification of Unbalanced Aberrations in the Genome of Equine Sarcoid Cells Using CGH Technique *Annals of Animal Science* (Vol. 16, pp. 79-85).

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

Calin, G. A., Dumitru, C. D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., Rassenti, L., Kipps, T., Negrini, M., Bullrich, F., & Croce, C. M. (2002). Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A, 99*(24), 15524-15529.

Calin, G. A., & Croce, C. M. (2006). MicroRNA signatures in human cancers. *Nat Rev Cancer, 6*(11), 857-866.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics, 10*, 421.

Campo, M. S. (2003). Papillomavirus and disease in humans and animals. *Vet Comp Oncol, 1*(1), 3-14.

Campo, M. S. (2006). Bovine Papillomavirus: Old Systems, New Lessons? In M. S. Campo (Ed.), *Papillomavirus research: from natural history to vaccines and beyond* (pp. 373-387). Norfolk, England (U.K.): Caister Academic Press.

Carr, E. A., Theon, A. P., Madewell, B. R., Griffey, S. M., & Hitchcock, M. E. (2001a). Bovine papillomavirus DNA in neoplastic and nonneoplastic tissues obtained from horses with and without sarcoids in the western United States. *Am J Vet Res, 62*(5), 741-744.

Carr, E. A., Theon, A. P., Madewell, B. R., Hitchcock, M. E., Schlegel, R., & Schiller, J. T. (2001b). Expression of a transforming gene (E5) of bovine papillomavirus in sarcoids obtained from horses. *Am J Vet Res, 62*(8), 1212-1217.

Castaldo, C., Di Meglio, F., Miraglia, R., Sacco, A. M., Romano, V., Bancone, C., Della Corte, A., Montagnani, S., & Nurzynska, D. (2013). Cardiac fibroblast-derived extracellular matrix (biomatrix) as a model for the studies of cardiac primitive cell biological properties in normal and pathological adult human heart. *Biomed Res Int, 2013*.

Chambers, G., Ellsmore, V. A., O'Brien, P. M., Reid, S. W., Love, S., Campo, M. S., & Nasir, L. (2003a). Sequence variants of bovine papillomavirus E5 detected in equine sarcoids. *Virus Res, 96*(1-2), 141-145.

Chambers, G., Ellsmore, V. A., O'Brien, P. M., Reid, S. W., Love, S., Campo, M. S., & Nasir, L. (2003b). Association of bovine papillomavirus with the equine sarcoid. *J Gen Virol, 84*(Pt 5), 1055-1062.

Chang, S. T., Thomas, M. J., Sova, P., Green, R. R., Palermo, R. E., & Katze, M. G. (2013). Next-Generation Sequencing of Small RNAs from HIV-Infected

Cells Identifies Phased microRNA Expression Patterns and Candidate Novel microRNAs Differentially Expressed upon Infection. *mBio, 4*(1).
Chen, C., Deng, B., Qiao, M., Zheng, R., Chai, J., Ding, Y., Peng, J., & Jiang, S. (2012). Solexa sequencing identification of conserved and novel microRNAs in backfat of Large White and Chinese Meishan pigs. *Plos One, 7*(2), e31426.
Christen, G., Gerber, V., Dolf, G., Burger, D., & Koch, C. (2014). Inheritance of equine sarcoid disease in Franches-Montagnes horses. *The Veterinary Journal, 199*(1), 68-71.
Cooper, B., Brimer, N., Stoler, M., & Vande Pol, S. B. (2006). Suprabasal overexpression of beta-1 integrin is induced by bovine papillomavirus type 1. *Virology, 355*(1), 102-114.
Corteggio, A., Urraro, C., Roperto, S., Roperto, F., & Borzacchiello, G. (2010). Phosphatidylinositol-3-kinase-AKT pathway, phospho-JUN and phospho-JNK expression in spontaneously arising bovine urinary bladder tumours. *J Comp Pathol, 143*(2-3), 173-178.
Corteggio, A., Di Geronimo, O., Roperto, S., Roperto, F., & Borzacchiello, G. (2011). Bovine papillomavirus E7 oncoprotein binds to p600 in naturally occurring equine sarcoids. *J Gen Virol, 92*(Pt 2), 378-382.
Corteggio, A., Di Geronimo, O., Roperto, S., Roperto, F., & Borzacchiello, G. (2012). Activated platelet-derived growth factor ß receptor and Ras-mitogen-activated protein kinase pathway in natural bovine urinary bladder carcinomas. *Vet J, 191*(3), 393-395.
Corteggio, A., Altamura, G., Roperto, F., & Borzacchiello, G. (2013). Bovine papillomavirus E5 and E7 oncoproteins in naturally occurring tumors: are two better than one? *Infectious Agents and Cancer, 8*, 1-1.
da Silveira, J. C., Veeramachaneni, D. N., Winger, Q. A., Carnevale, E. M., & Bouma, G. J. (2012). Cell-secreted vesicles in equine ovarian follicular fluid contain miRNAs and proteins: a possible new form of cell communication within the ovarian follicle. *Biol Reprod, 86*(3), 71.
da Silveira, J. C., Carnevale, E. M., Winger, Q. A., & Bouma, G. J. (2014). Regulation of ACVR1 and ID2 by cell-secreted exosomes during follicle maturation in the mare. *Reprod Biol Endocrinol, 12*, 44.
Dareng, E. O., Ma, B., Famooto, A. O., Akarolo-Anthony, S. N., Offiong, R. A., Olaniyan, O., Dakum, P. S., Wheeler, C. M., Fadrosh, D., Yang, H., Gajer, P., Brotman, R. M., Ravel, J., & Adebamowo, C. A. (2016). Prevalent high-risk HPV infection and vaginal microbiota in Nigerian women. *Epidemiology and Infection, 144*(1), 123-137.
Das, A. V., & Pillai, R. M. (2015). Implications of miR cluster 143/145 as universal anti-oncomiRs and their dysregulation during tumorigenesis. *Cancer cell international, 15*(1), 1.
Das, P. J., McCarthy, F., Vishnoi, M., Paria, N., Gresham, C., Li, G., Kachroo, P., Sudderth, A. K., Teague, S., Love, C. C., Varner, D. D., Chowdhary, B. P., & Raudsepp, T. (2013). Stallion Sperm Transcriptome Comprises Functionally Coherent Coding and Regulatory RNAs as Revealed by Microarray Analysis and RNA-seq. *Plos One, 8*(2).
Davenport, M., Poles, J., Leung, J. M., Wolff, M. J., Abidi, W. M., Ullman, T., Mayer, L., Cho, I., & Loke, P. n. (2014). Metabolic alterations to the mucosal microbiota in inflammatory bowel disease. *Inflammatory bowel diseases, 20*(4), 723.

de Freitas, A. C., Coimbra, E. C., & Leitao Mda, C. (2014). Molecular targets of HPV oncoproteins: potential biomarkers for cervical carcinogenesis. *Biochim Biophys Acta, 1845*(2), 91-103.
de Martel, C., Ferlay, J., Franceschi, S., Vignat, J., Bray, F., Forman, D., & Plummer, M. (2012). Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *The Lancet Oncology, 13*(6), 607-615.
de Villiers, E.-M., Fauquet, C., Broker, T. R., Bernard, H.-U., & zur Hausen, H. (2004). Classification of papillomaviruses. *Virology, 324*(1), 17-27.
DeMasi, J., Huh, K. W., Nakatani, Y., Munger, K., & Howley, P. M. (2005). Bovine papillomavirus E7 transformation function correlates with cellular p600 protein binding. *Proc Natl Acad Sci U S A, 102*(32), 11486-11491.
DeMasi, J., Chao, M. C., Kumar, A. S., & Howley, P. M. (2007). Bovine papillomavirus E7 oncoprotein inhibits anoikis. *Journal of Virology, 81*(17), 9419-9425.
Desjardin, C., Vaiman, A., Mata, X., Legendre, R., Laubier, J., Kennedy, S. P., Laloe, D., Barrey, E., Jacques, C., Cribiu, E. P., & Schibler, L. (2014). Next-generation sequencing identifies equine cartilage and subchondral bone miRNAs and suggests their involvement in osteochondrosis physiopathology. *BMC Genomics, 15*(1), 798.
Diaz-Gonzalez, S., M., Deas, J., Benitez-Boijseauneau, O., Gomez-Ceron, C., Bermudez-Morales, V. H., Rodriguez-Dorantes, M., Perez-Plasencia, C., & Peralta-Zaragoza, O. (2015). Utility of microRNAs and siRNAs in cervical carcinogenesis. *Biomed Res Int, 2015*, 374924.
DiMaio, D., & Petti, L. M. (2013). The E5 proteins. *Virology, 445*(1), 99-114.
Donadeu, F. X., & Schauer, S. N. (2013). Differential miRNA expression between equine ovulatory and anovulatory follicles. *Domest Anim Endocrinol, 45*(3), 122-125.
Doorbar, J., Quint, W., Banks, L., Bravo, I. G., Stoler, M., Broker, T. R., & Stanley, M. A. (2012). The biology and life-cycle of human papillomaviruses. *Vaccine, 30 Suppl 5*, F55-70.
Doorbar, J., Egawa, N., Griffin, H., Kranjec, C., & Murakami, I. (2015). Human papillomavirus molecular biology and disease association. *Rev Med Virol, 25 Suppl 1*, 2-23.
Fang, R., Xiao, T., Fang, Z., Sun, Y., Li, F., Gao, Y., Feng, Y., Li, L., Wang, Y., & Liu, X. (2012). MicroRNA-143 (miR-143) regulates cancer glycolysis via targeting hexokinase 2 gene. *Journal of Biological Chemistry, 287*(27), 23227-23235.
Fei, J., Li, Y., Zhu, X., & Luo, X. (2012). miR-181a post-transcriptionally downregulates oncogenic RalA and contributes to growth inhibition and apoptosis in chronic myelogenous leukemia (CML). *PLoS One, 7*(3), e32834.
Filipowicz, W., Bhattacharyya, S. N., & Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet, 9*(2), 102-114.
Finlay, M., Yuan, Z., Burden, F., Trawford, A., Morgan, I. M., Campo, M. S., & Nasir, L. (2009). The detection of Bovine Papillomavirus type 1 DNA in flies. *Virus Res, 144*(1-2), 315-317.
Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W., & Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Research, 40*(1), 37-52.

Friedman, R. C., Farh, K. K.-H., Burge, C. B., & Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research, 19*(1), 92-105.
Gao, W., Yu, Y., Cao, H., Shen, H., Li, X., Pan, S., & Shu, Y. (2010). Deregulated expression of miR-21, miR-143 and miR-181a in non small cell lung cancer is related to clinicopathologic characteristics or patient prognosis. *Biomed Pharmacother, 64*(6), 399-408.
Garcia-Lora, A., Algarra, I., & Garrido, F. (2003). MHC class I antigens, immune surveillance, and tumor immune escape. *J Cell Physiol, 195*(3), 346-355.
Gaynor, A. M., Zhu, K. W., Cruz, F. N. D., Affolter, V. K., & Pesavento, P. A. (2016). Localization of Bovine Papillomavirus Nucleic Acid in Equine Sarcoids. *Vet Pathol, 53*(3), 567-573.
Gerber, H. (1989). Sir Frederick Hobday memorial lecture. The genetic basis of some equine diseases. *Equine Vet J, 21*(4), 244-248.
Gil da Costa, R. M., & Medeiros, R. (2014). Bovine papillomavirus: opening new trends for comparative pathology. *Arch Virol, 159*(2), 191-198.
Giza, D. E., Vasilescu, C., & Calin, G. A. (2014). Key principles of miRNA involvement in human diseases. *Discoveries, 2*(4), e34.
Gocze, K., Gombos, K., Juhasz, K., Kovacs, K., Kajtar, B., Benczik, M., Gocze, P., Patczai, B., Arany, I., & Ember, I. (2013). Unique microRNA expression profiles in cervical cancer. *Anticancer Res, 33*(6), 2561-2567.
Goldstein, D. J., Kulke, R., Dimaio, D., & Schlegel, R. (1992). A glutamine residue in the membrane-associating domain of the bovine papillomavirus type 1 E5 oncoprotein mediates its binding to a transmembrane component of the vacuolar H(+)-ATPase. *J Virol, 66*(1), 405-413.
Gomez-Gomez, Y., Organista-Nava, J., & Gariglio, P. (2013). Deregulation of the miRNAs expression in cervical cancer: human papillomavirus implications. *Biomed Res Int, 2013*, 407052.
Gómez-Gómez, Y., Organista-Nava, J., & Gariglio, P. (2013). Deregulation of the miRNAs expression in cervical cancer: Human papillomavirus implications. *Biomed Res Int, 2013*.
Goodrich, L., Gerber, H., Marti, E., & Antczak, D. F. (1998). Equine sarcoids. *Vet Clin North Am Equine Pract, 14*(3), 607-623, vii.
Goto, Y., Kojima, S., Nishikawa, R., Enokida, H., Chiyomaru, T., Kinoshita, T., Nakagawa, M., Naya, Y., Ichikawa, T., & Seki, N. (2014). The microRNA-23b/27b/24-1 cluster is a disease progression marker and tumor suppressor in prostate cancer. *Oncotarget, 5*(17), 7748-7759.
Greco, D., Kivi, N., Qian, K., Leivonen, S. K., Auvinen, P., & Auvinen, E. (2011). Human papillomavirus 16 E5 modulates the expression of host microRNAs. *Plos One, 6*(7), e21646.
Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., & Enright, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research, 34*(suppl 1), D140-D144.
Griffiths-Jones, S., Hui, J. H., Marco, A., & Ronshaugen, M. (2011). MicroRNA evolution by arm switching. *EMBO Rep, 12*(2), 172-177.
Gunaratne, P. H., Coarfa, C., Soibam, B., & Tandon, A. (2012). miRNA data analysis: next-gen sequencing. *Next-Generation MicroRNA Expression Profiling Technology: Methods and Protocols*, 273-288.
Guo, L., & Chen, F. (2014). A challenge for miRNA: multiple isomiRs in miRNAomics. *Gene, 544*(1), 1-7.
Guo, X. K., Zhang, Q., Gao, L., Li, N., Chen, X. X., & Feng, W. H. (2013). Increasing expression of microRNA 181 inhibits porcine reproductive and

respiratory syndrome virus replication and has implications for controlling virus infection. *J Virol, 87*(2), 1159-1171.
Gupta, A., Swaminathan, G., Martin-Garcia, J., & Navas-Martin, S. (2012). MicroRNAs, hepatitis C virus, and HCV/HIV-1 co-infection: new insights in pathogenesis and therapy. *Viruses, 4*(11), 2485-2513.
Hamann, J., & Grabner, A. (2005). Equine Sarcoid: the most-common skin tumour of horses. *PFERDEHEILKUNDE, 21*(4), 273-279.
Han, X., Chen, Y., Yao, N., Liu, H., & Wang, Z. (2015). MicroRNA let-7b suppresses human gastric cancer malignancy by targeting ING1. *Cancer gene therapy, 22*(3), 122-129.
Haralambus, R., Burgstaller, J., Klukowska-Rötzler, J., Steinborn, R., Buchinger, S., Gerber, V., & Brandt, S. (2010). Intralesional bovine papillomavirus DNA loads reflect severity of equine sarcoid disease. *Equine Vet J, 42*(4), 327-331.
Hartl, B., Hainisch, E. K., Shafti-Keramat, S., Kirnbauer, R., Corteggio, A., Borzacchiello, G., Tober, R., Kainzbauer, C., Pratscher, B., & Brandt, S. (2011). Inoculation of young horses with bovine papillomavirus type 1 virions leads to early infection of PBMCs prior to pseudo-sarcoid formation. *J Gen Virol, 92*(Pt 10), 2437-2445.
Heneghan, H. M., Miller, N., & Kerin, M. J. (2010). MiRNAs as biomarkers and therapeutic targets in cancer. *Current opinion in pharmacology, 10*(5), 543-550.
Hermeking, H. (2007). p53 enters the microRNA world. *Cancer Cell, 12*(5), 414-418.
Hertel, J., Bartschat, S., Wintsche, A., Otto, C., of the Bioinformatics Computer Lab, T. S., & Stadler, P. F. (2012). Evolution of the let-7 microRNA Family. *RNA Biology, 9*(3), 231-241.
Hill, A. B. (1965). The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine, 58*(5), 295-300.
Honegger, A., Schilling, D., Bastian, S., Sponagel, J., Kuryshev, V., Sultmann, H., Scheffner, M., Hoppe-Seyler, K., & Hoppe-Seyler, F. (2015). Dependence of intracellular and exosomal microRNAs on viral E6/E7 oncogene expression in HPV-positive tumor cells. *PLoS Pathog, 11*(3), e1004712.
Horwitz, B. H., Burkhardt, A. L., Schlegel, R., & DiMaio, D. (1988). 44-amino-acid E5 transforming protein of bovine papillomavirus requires a hydrophobic core and specific carboxyl-terminal amino acids. *Mol Cell Biol, 8*(10), 4071-4078.
Huang, E., Liu, R., & Chu, Y. (2015). miRNA-15a/16: as tumor suppressors and more. *Future Oncol, 11*(16), 2351-2363.
Hung, C. H., Hu, T. H., Lu, S. N., Kuo, F. Y., Chen, C. H., Wang, J. H., Huang, C. M., Lee, C. M., Lin, C. Y., Yen, Y. H., & Chiu, Y. C. (2015). Circulating microRNAs as biomarkers for diagnosis of early hepatocellular carcinoma associated with hepatitis B virus. *Int J Cancer*.
Hung, C. H., Hu, T. H., Lu, S. N., Kuo, F. Y., Chen, C. H., Wang, J. H., Huang, C. M., Lee, C. M., Lin, C. Y., Yen, Y. H., & Chiu, Y. C. (2016). Circulating microRNAs as biomarkers for diagnosis of early hepatocellular carcinoma associated with hepatitis B virus. *Int J Cancer, 138*(3), 714-720.
Hunninghake, G. W., Costabel, U., Ando, M., Baughman, R., Cordier, J. F., du Bois, R., Eklund, A., Kitaichi, M., Lynch, J., Rizzato, G., Rose, C., Selroos, O., Semenzato, G., & Sharma, O. P. (1999). ATS/ERS/WASOG statement on sarcoidosis. American Thoracic Society/European Respiratory

Society/World Association of Sarcoidosis and other Granulomatous Disorders. *Sarcoidosis Vasc Diffuse Lung Dis, 16*(2), 149-173.
Iaconetti, C., De Rosa, S., Polimeni, A., Sorrentino, S., Gareri, C., Carino, A., Sabatino, J., Colangelo, M., Curcio, A., & Indolfi, C. (2015). Down-regulation of miR-23b induces phenotypic switching of vascular smooth muscle cells in vitro and in vivo. *Cardiovasc Res*.
Iorio, M. V., & Croce, C. M. (2012). MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. *EMBO molecular medicine, 4*(3), 143-159.
Jackson, C. (1936). *The incidence and pathology of tumours of domesticated animals in South Africa: A study of the Onderstepoort collection of neoplasms with special reference to their histopathology* (Vol. 6): Government Printer, South Africa.
Jagadeeswaran, G., Zheng, Y., Sumathipala, N., Jiang, H., Arrese, E. L., Soulages, J. L., Zhang, W., & Sunkar, R. (2010). Deep sequencing of small RNA libraries reveals dynamic regulation of conserved and novel microRNAs and microRNA-stars during silkworm development. *BMC Genomics, 11*, 52.
Jandova, V., Klukowska-Rotzler, J., Dolf, G., Janda, J., Roosje, P., Marti, E., Koch, C., Gerber, V., & Swinburne, J. (2012). Whole genome scan identifies several chromosomal regions linked to equine sarcoids. *Schweiz Arch Tierheilkd, 154*(1), 19-25.
Jansson, M. D., & Lund, A. H. (2012). MicroRNA and cancer. *Molecular oncology, 6*(6), 590-610.
Ji, Z., Wang, G., Xie, Z., Wang, J., Zhang, C., Dong, F., & Chen, C. (2012). Identification of novel and differentially expressed MicroRNAs of dairy goat mammary gland tissues using solexa sequencing and bioinformatics. *Plos One, 7*(11), e49463.
Johnson, S. M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K. L., Brown, D., & Slack, F. J. (2005). RAS is regulated by the let-7 microRNA family. *Cell, 120*(5), 635-647.
Joyce, J. G., Tung, J. S., Przysiecki, C. T., Cook, J. C., Lehman, E. D., Sands, J. A., Jansen, K. U., & Keller, P. M. (1999). The L1 major capsid protein of human papillomavirus type 11 recombinant virus-like particles interacts with heparin and cell-surface glycosaminoglycans on human keratinocytes. *J Biol Chem, 274*(9), 5810-5822.
Jung, H. M., Phillips, B. L., & Chan, E. K. (2014). miR-375 activates p21 and suppresses telomerase activity by coordinately regulating HPV E6/E7, E6AP, CIP2A, and 14-3-3ζ. *Mol Cancer, 13*(1), 1.
Kainzbauer, C., Rushton, J., Tober, R., Scase, T., Nell, B., Sykora, S., & Brandt, S. (2012). Bovine papillomavirus type 1 and Equus caballus papillomavirus 2 in equine squamous cell carcinoma of the head and neck in a Connemara mare. *Equine Vet J, 44*(1), 112-115.
Kamper, N., Day, P. M., Nowak, T., Selinka, H. C., Florin, L., Bolscher, J., Hilbig, L., Schiller, J. T., & Sapp, M. (2006). A membrane-destabilizing peptide in capsid protein L2 is required for egress of papillomavirus genomes from endosomes. *J Virol, 80*(2), 759-768.
Kanellopoulou, C., Muljo, S. A., Kung, A. L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D. M., & Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev, 19*(4), 489-501.

Kemp-Symonds, J. (2000). The detection and sequencing of bovine papillomavirus type 1 and 2 DNA from Musca autumnalis (Diptera: Muscidae) face flies infesting sarcoid-affected horses. *Royal Veterinary College, London, UK*.

Kent, O. A., McCall, M. N., Cornish, T. C., & Halushka, M. K. (2014). Lessons from miR-143/145: the importance of cell-type localization of miRNAs. *Nucleic Acids Research, 42*(12), 7528-7538.

Kidney, B. A., & Berrocal, A. (2008). Sarcoids in two captive tapirs (Tapirus bairdii): clinical, pathological and molecular study. *Vet Dermatol, 19*(6), 380-384.

Kilcoyne, I., Watson, J. L., Kass, P. H., & Spier, S. J. (2013). Incidence, management, and outcome of complications of castration in equids: 324 cases (1998-2008). *J Am Vet Med Assoc, 242*(6), 820-825.

Kim, C. H., Kim, H. K., Rettig, R. L., Kim, J., Lee, E. T., Aprelikova, O., Choi, I. J., Munroe, D. J., & Green, J. E. (2011a). miRNA signature associated with outcome of gastric cancer patients following chemotherapy. *BMC Med Genomics, 4*, 79.

Kim, M. C., Lee, S. W., Ryu, D. Y., Cui, F. J., Bhak, J., & Kim, Y. (2014). Identification and characterization of microRNAs in normal equine tissues by Next Generation Sequencing. *Plos One, 9*(4), e93662.

Kim, N. H., Kim, H. S., Li, X. Y., Lee, I., Choi, H. S., Kang, S. E., Cha, S. Y., Ryu, J. K., Yoon, D., Fearon, E. R., Rowe, R. G., Lee, S., Maher, C. A., Weiss, S. J., & Yook, J. I. (2011b). A p53/miRNA-34 axis regulates Snail1-dependent cancer cell epithelial-mesenchymal transition. *J Cell Biol, 195*(3), 417-433.

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

Klein, O., Kegler-Ebo, D., Su, J., Smith, S., & DiMaio, D. (1999). The bovine papillomavirus E5 protein requires a juxtamembrane negative charge for activation of the platelet-derived growth factor beta receptor and transformation of C127 cells. *J Virol, 73*(4), 3264-3272.

Knottenbelt, D., Edwards, S., & Daniel, E. (1995). Diagnosis and treatment of the equine sarcoid. *In Practice, 17*(3), 123-129.

Knottenbelt, D. (2005a). Equine sarcoid *Pascoe's Principles and Practice of Equine Dermatology 2nd Edition,*, 387-407.

Knottenbelt, D. (2005b). A suggested clinical classification for the equine sarcoid *Clinical Techniques in Equine Practice, 4, 278-295*.

Knottenbelt, D. C., & Kelly, D. F. (2000). The diagnosis and treatment of periorbital sarcoid in the horse: 445 cases from 1974 to 1999. *Vet Ophthalmol, 3*(2-3), 169-191.

Kozomara, A., & Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research, 42*(D1), D68-D73.

Lajer, C. B., Nielsen, F. C., Friis-Hansen, L., Norrild, B., Borup, R., Garnaes, E., Rossing, M., Specht, L., Therkildsen, M. H., Nauntofte, B., Dabelsteen, S., & von Buchwald, C. (2011). Different miRNA signatures of oral and pharyngeal squamous cell carcinomas: a prospective translational study. *Br J Cancer, 104*(5), 830-840.

Lajer, C. B., Garnaes, E., Friis-Hansen, L., Norrild, B., Therkildsen, M. H., Glud, M., Rossing, M., Lajer, H., Svane, D., Skotte, L., Specht, L., Buchwald, C., & Nielsen, F. C. (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.
Lancaster, W. D., Olson, C., & Meinke, W. (1977). Bovine papilloma virus: presence of virus-specific DNA sequences in naturally occurring equine tumors. *Proceedings of the National Academy of Sciences, 74*(2), 524-528.
Lange, C. E., Vetsch, E., Ackermann, M., Favrot, C., & Tobler, K. (2013). Four novel papillomavirus sequences support a broad diversity among equine papillomaviruses. *Journal of general virology, 94*(6), 1365-1372.
Lazary, S., Gerber, H., Glatt, P. A., & Straub, R. (1985). Equine leucocyte antigens in sarcoid-affected horses. *Equine Vet J, 17*(4), 283-286.
Lee, C. T., Risom, T., & Strauss, W. M. (2007). Evolutionary conservation of microRNA regulatory circuits: an examination of microRNA gene complexity and conserved microRNA-target interactions through metazoan phylogeny. *DNA Cell Biol, 26*(4), 209-218.
Lee, H., Han, S., Kwon, C. S., & Lee, D. (2016). Biogenesis and regulation of the let-7 miRNAs and their functional implications. *Protein Cell, 7*(2), 100-113.
Lee, R. C., Feinbaum, R. L., & 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. H., Lee, C. R., Rigas, N. K., Kim, R. H., Kang, M. K., Park, N. H., & Shin, K. H. (2015). Human papillomavirus 16 (HPV16) enhances tumor growth and cancer stemness of HPV-negative oral/oropharyngeal squamous cell carcinoma cells via miR-181 regulation. *Papillomavirus Res, 1*, 116-125.
Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H., & Kim, V. N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *EMBO J, 23*(20), 4051-4060.
Lewis, B. P., Burge, C. B., & 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.
Li, B., Hu, Y., Ye, F., Li, Y., Lv, W., & Xie, X. (2010). Reduced miR-34a expression in normal cervical tissues and cervical lesions with high-risk human papillomavirus infection. *International Journal of Gynecological Cancer, 20*(4), 597-604.
Li, S., Meng, H., Zhou, F., Zhai, L., Zhang, L., Gu, F., Fan, Y., Lang, R., Fu, L., Gu, L., & Qi, L. (2013). MicroRNA-132 is frequently down-regulated in ductal carcinoma in situ (DCIS) of breast and acts as a tumor suppressor by inhibiting cell proliferation. *Pathol Res Pract, 209*(3), 179-183.
Li, Y., Wang, F., Xu, J., Ye, F., Shen, Y., Zhou, J., Lu, W., Wan, X., Ma, D., & Xie, X. (2011). Progressive miRNA expression profiles in cervical carcinogenesis and identification of HPV-related target genes for miR-29. *J Pathol, 224*(4), 484-495.
Liu, C. (2013). The role of microRNAs in tumors. *Archives of pharmacal research, 36*(10), 1169-1177.
Liu, L., Yu, X., Guo, X., Tian, Z., Su, M., Long, Y., Huang, C., Zhou, F., Liu, M., & Wu, X. (2012). miR-143 is downregulated in cervical cancer and promotes apoptosis and inhibits tumor formation by targeting Bcl-2. *Mol Med Rep, 5*(3), 753-760.
Liu, X., Yu, H., Cai, H., & Wang, Y. (2014). The expression and clinical significance of miR-132 in gastric cancer patients. *Diagn Pathol, 9*, 57.

Liu, Y., Hong, Y., Androphy, E. J., & Chen, J. J. (2000). Rb-independent induction of apoptosis by bovine papillomavirus type 1 E7 in response to tumor necrosis factor alpha. *J Biol Chem, 275*(40), 30894-30900.

Lory, S., von Tscharner, C., Marti, E., Bestetti, G., Grimm, S., & Waldvogel, A. (1993). In situ hybridisation of equine sarcoids with bovine papilloma virus. *Vet Rec, 132*(6), 132-133.

Lotterman, C. D., Kent, O. A., & Mendell, J. T. (2008). Functional integration of microRNAs into oncogenic and tumor suppressor pathways. *Cell Cycle, 7*(16), 2493-2499.

Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B. L., Mak, R. H., Ferrando, A. A., Downing, J. R., Jacks, T., Horvitz, H. R., & Golub, T. R. (2005). MicroRNA expression profiles classify human cancers. *Nature, 435*(7043), 834-838.

Lu, Y.-C., Chang, J. T., Liao, C.-T., Kang, C.-J., Huang, S.-F., Chen, I. H., Huang, C.-C., Huang, Y.-C., Chen, W.-H., Tsai, C.-Y., Wang, H.-M., Yen, T.-C., You, G.-R., Chiang, C.-H., & Cheng, A.-J. (2014). OncomiR-196 promotes an invasive phenotype in oral cancer through the NME4-JNK-TIMP1-MMP signaling pathway. *Mol Cancer, 13*, 218.

Lunardi, M., Alfieri, A. A., Otonel, R. A., de Alcantara, B. K., Rodrigues, W. B., de Miranda, A. B., & Alfieri, A. F. (2013a). Genetic characterization of a novel bovine papillomavirus member of the Deltapapillomavirus genus. *Vet Microbiol, 162*(1), 207-213.

Lunardi, M., de Alcantara, B. K., Otonel, R. A., Rodrigues, W. B., Alfieri, A. F., & Alfieri, A. A. (2013b). Bovine papillomavirus type 13 DNA in equine sarcoids. *J Clin Microbiol, 51*(7), 2167-2171.

Ma, G., Dai, W., Sang, A., Yang, X., & Gao, C. (2014). Upregulation of microRNA-23a/b promotes tumor progression and confers poor prognosis in patients with gastric cancer. *Int J Clin Exp Pathol, 7*(12), 8833-8840.

Maher, C., Stein, L., & Ware, D. (2006). Evolution of Arabidopsis microRNA families through duplication events. *Genome Research, 16*(4), 510-519.

Mahlmann, K., Hamza, E., Marti, E., Dolf, G., Klukowska, J., Gerber, V., & Koch, C. (2014). Increased FOXP3 expression in tumour-associated tissues of horses affected with equine sarcoid disease. *Vet J, 202*(3), 516-521.

Maiolino, P., Ozkul, A., Sepici-Dincel, A., Roperto, F., Yucel, G., Russo, V., Urraro, C., Luca, R., Riccardi, M. G., Martano, M., Borzacchiello, G., Esposito, I., & Roperto, S. (2013). Bovine papillomavirus type 2 infection and microscopic patterns of urothelial tumors of the urinary bladder in water buffaloes. *Biomed Res Int, 2013*, 937918.

Manzano, M., Forte, E., Raja, A. N., Schipma, M. J., & Gottwein, E. (2015). Divergent target recognition by coexpressed 5′-isomiRs of miR-142-3p and selective viral mimicry. *RNA, 21*(9), 1606-1620.

Marais, H. J., Nel, P., Bertschinger, H. J., Schoeman, J. P., & Zimmerman, D. (2007). Prevalence and body distribution of sarcoids in South African Cape mountain zebra (Equus zebra zebra). *J S Afr Vet Assoc, 78*(3), 145-148.

Marchetti, B., Ashrafi, G. H., Tsirimonaki, E., O'Brien, P. M., & Campo, M. S. (2002). The bovine papillomavirus oncoprotein E5 retains MHC class I molecules in the Golgi apparatus and prevents their transport to the cell surface. *Oncogene, 21*(51), 7808-7816.

Marchetti, B., Gault, E. A., Cortese, M. S., Yuan, Z., Ellis, S. A., Nasir, L., & Campo, M. S. (2009). Bovine papillomavirus type 1 oncoprotein E5 inhibits equine MHC class I and interacts with equine MHC I heavy chain. *J Gen Virol, 90*(Pt 12), 2865-2870.

Marco, A., Ninova, M., Ronshaugen, M., & Griffiths-Jones, S. (2013). Clusters of microRNAs emerge by new hairpins in existing transcripts. *Nucleic Acids Research*, gkt534.
Martens, A., De Moor, A., Demeulemeester, J., & Ducatelle, R. (2000). Histopathological characteristics of five clinical types of equine sarcoid. *Res Vet Sci, 69*(3), 295-300.
Martens, A., De Moor, A., Demeulemeester, J., & Peelman, L. (2001a). Polymerase chain reaction analysis of the surgical margins of equine sarcoids for bovine papilloma virus DNA. *Vet Surg, 30*(5), 460-467.
Martens, A., De Moor, A., Vlaminck, L., Pile, F., & Steenhaut, M. (2001b). Evaluation of excision, cryosurgery and local BCG vaccination for the treatment of equine sarcoids. *Veterinary Record, 149*(22), 665-669.
Martens, A., deMoor, A., & Ducatelle, R. (2001c). PCR detection of bovine papillomavirus DNA in superficial swabs and scrapings from equine sarcoids The Veterinary Journal (Vol. 161, pp. 280-286.).
Marti, E., Lazary, S., Antczak, D. F., & Gerber, H. (1993). Report of the first international workshop on equine sarcoid. *Equine Vet J, 25*(5), 397-407.
Martinez, I., Gardiner, A., Board, K., Monzon, F., Edwards, R., & Khan, S. (2008). Human papillomavirus type 16 reduces the expression of microRNA-218 in cervical carcinoma cells. *Oncogene, 27*(18), 2575-2582.
Masliah-Planchon, J., Garinet, S., & Pasmant, E. (2015). RAS-MAPK pathway epigenetic activation in cancer: miRNAs in action. *Oncotarget*.
McBee, W. C., Gardiner, A. S., Edwards, R. P., Lesnock, J. L., Bhargava, R., Austin, R. M., Guido, R. S., & Khan, S. A. (2011). MicroRNA analysis in human papillomavirus (HPV)-associated cervical neoplasia and cancer. *Journal of Carcinogenesis & Mutagenesis, 2*(1).
McKenna, D. J., McDade, S. S., Patel, D., & McCance, D. J. (2010). MicroRNA 203 expression in keratinocytes is dependent on regulation of p53 levels by E6. *Journal of Virology, 84*(20), 10644-10652.
McMillan, N. A., Payne, E., Frazer, I. H., & Evander, M. (1999). Expression of the alpha6 integrin confers papillomavirus binding upon receptor-negative B-cells. *Virology, 261*(2), 271-279.
Melar-New, M., & Laimins, L. A. (2010). Human papillomaviruses modulate expression of microRNA 203 upon epithelial differentiation to control levels of p63 proteins. *Journal of Virology, 84*(10), 5212-5221.
Melton, C., & Blelloch, R. (2010). MicroRNA Regulation of Embryonic Stem Cell Self-Renewal and Differentiation. *Adv Exp Med Biol, 695*, 105-117.
Meredith, D., Elser, A. H., Wolf, B., Soma, L. R., Donawick, W. J., & Lazary, S. (1986). Equine leukocyte antigens: relationships with sarcoid tumors and laminitis in two pure breeds. *Immunogenetics, 23*(4), 221-225.
Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat Rev Genet, 11*(1), 31-46.
Miller, R. I., & Campbell, R. S. (1982). A survey of granulomatous and neoplastic diseases of equine skin in north Queensland. *Aust Vet J, 59*(2), 33-37.
Mo, J.-S., Alam, K. J., Kang, I.-H., Park, W. C., Seo, G.-S., Choi, S.-C., Kim, H.-S., Moon, H.-B., Yun, K.-J., & Chae, S.-C. (2015). MicroRNA 196B regulates FAS-mediated apoptosis in colorectal cancer cells. *Oncotarget, 6*(5), 2843-2855.
Mohammed, H. O., Rebhun, W. C., & Antczak, D. F. (1992). Factors associated with the risk of developing sarcoid tumours in horses. *Equine Vet J, 24*(3), 165-168.

Moody, C. A., & 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., Hirst, M., Eaves, C. J., & Marra, M. A. (2008). Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Research, 18*(4), 610-621.
Morozova, O., & Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics, 92*(5), 255-264.
Mosseri, S., Hetzel, U., Hahn, S., Michaloupoulou, E., Sallabank, H. C., Knottenbelt, D. C., & Kipar, A. (2014). Equine sarcoid: In situ demonstration of matrix metalloproteinase expression. *Vet J, 202*(2), 279-285.
Munday, J. S. (2014). Bovine and human papillomaviruses: a comparative review. *Vet Pathol, 51*(6), 1063-1075.
Münger, K., Baldwin, A., Edwards, K. M., Hayakawa, H., Nguyen, C. L., Owens, M., Grace, M., & Huh, K. (2004). Mechanisms of Human Papillomavirus-Induced Oncogenesis. *Journal of Virology, 78*(21), 11451-11460.
Myklebust, M., Bruland, O., Fluge, Ø., Skarstein, A., Balteskard, L., & Dahl, O. (2011). MicroRNA-15b is induced with E2F-controlled genes in HPV-related cancer. *Br J Cancer, 105*(11), 1719-1725.
Nagaraj, A. B., Joseph, P., & DiFeo, A. (2015). miRNAs as prognostic and therapeutic tools in epithelial ovarian cancer. *Biomark Med, 9*(3), 241-257.
Nasir, L., McFarlane, S., & Reid, S. (1999). Mutational status of the tumour suppressor gene (p53) in donkey sarcoid tumours. *The Veterinary Journal, 157*(1), 99-101.
Nasir, L., & Reid, S. W. (1999). Bovine papillomaviral gene expression in equine sarcoid tumours. *Virus Res, 61*(2), 171-175.
Nasir, L., Gault, E., Morgan, I. M., Chambers, G., Ellsmore, V., & Campo, M. S. (2007). Identification and functional analysis of sequence variants in the long control region and the E2 open reading frame of bovine papillomavirus type 1 isolated from equine sarcoids. *Virology, 364*(2), 355-361.
Nasir, L., & Campo, M. S. (2008). Bovine papillomaviruses: their role in the aetiology of cutaneous tumours of bovids and equids. *Vet Dermatol, 19*(5), 243-254.
Nasir, L., & Brandt, S. (2013). Papillomavirus associated diseases of the horse. *Vet Microbiol, 17*, 159-167.
NEHS. (2014). National Equine Health Survey (NEHS) *AHT / BEVA / DEFRA Equine Quarterly Disease Surveillance Report, 10*(1).
NEHS. (2015, 22nd February 2016). National Equine Health Survey (NEHS) Retrieved from https://www.bluecross.org.uk/nehs-2015-results
Neilsen, C. T., Goodall, G. J., & Bracken, C. P. (2012). IsomiRs–the overlooked repertoire in the dynamic microRNAome. *Trends in Genetics, 28*(11), 544-549.
Nel, P. J., Bertschinger, H., Williams, J., & Thompson, P. N. (2006). Descriptive study of an outbreak of equine sarcoid in a population of Cape mountain zebra (Equus zebra zebra) in the Gariep Nature Reserve. *J S Afr Vet Assoc, 77*(4), 184-190.

Olson, C., Jr., & Cook, R. H. (1951). Cutaneous sarcoma-like lesions of the horse caused by the agent of bovine papilloma. *Proc Soc Exp Biol Med, 77*(2), 281-284.

Otten, N., von Tscharner, C., Lazary, S., Antczak, D. F., & Gerber, H. (1993). DNA of bovine papillomavirus type 1 and 2 in equine sarcoids: PCR detection and direct sequencing. *Arch Virol, 132*(1-2), 121-131.

Pagliuca, A., Valvo, C., Fabrizi, E., Di Martino, S., Biffoni, M., Runci, D., Forte, S., De Maria, R., & Ricci-Vitiani, L. (2013). Analysis of the combined action of miR-143 and miR-145 on oncogenic pathways in colorectal cancer cells reveals a coordinate program of gene repression. *Oncogene, 32*(40), 4806-4813.

Pangty, K., Singh, S., Goswami, R., Saikumar, G., & Somvanshi, R. (2010). Detection of BPV-1 and -2 and quantification of BPV-1 by real-time PCR in cutaneous warts in cattle and buffaloes. *Transbound Emerg Dis, 57*(3), 185-196.

Park, P., Copeland, W., Yang, L., Wang, T., Botchan, M. R., & Mohr, I. J. (1994). The cellular DNA polymerase alpha-primase is required for papillomavirus DNA replication and associates with the viral E1 helicase. *Proceedings of the National Academy of Sciences of the United States of America, 91*(18), 8700-8704.

Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., Hayward, D. C., Ball, E. E., Degnan, B., Muller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., & Ruvkun, G. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature, 408*(6808), 86-89.

Pereira, P. M., Marques, J. P., Soares, A. R., Carreto, L., & Santos, M. A. (2010). MicroRNA expression variability in human cervical tissues. *Plos One, 5*(7), e11780.

Petti, L., Nilson, L. A., & DiMaio, D. (1991). Activation of the platelet-derived growth factor receptor by the bovine papillomavirus E5 transforming protein. *EMBO J, 10*(4), 845-855.

Petti, L. M., & Ray, F. A. (2000). Transformation of mortal human fibroblasts and activation of a growth inhibitory pathway by the bovine papillomavirus E5 oncoprotein. *Cell Growth Differ, 11*(7), 395-408.

Platt, R. N., 2nd, Vandewege, M. W., Kern, C., Schmidt, C. J., Hoffmann, F. G., & Ray, D. A. (2014). Large numbers of novel miRNAs originate from DNA transposons and are coincident with a large species radiation in bats. *Mol Biol Evol, 31*(6), 1536-1545.

Pouladi, N., Kouhsari, S. M., Feizi, M. H., Gavgani, R. R., & Azarfam, P. (2013). Overlapping region of p53/wrap53 transcripts: mutational analysis and sequence similarity with microRNA-4732-5p. *Asian Pac J Cancer Prev, 14*(6), 3503-3507.

Pritchard, C. C., Cheng, H. H., & Tewari, M. (2012). MicroRNA profiling: approaches and considerations. *Nature reviews genetics, 13*(5), 358-369.

Qian, K., Pietila, T., Ronty, M., Michon, F., Frilander, M. J., Ritari, J., Tarkkanen, J., Paulin, L., Auvinen, P., & Auvinen, E. (2013). Identification and validation of human papillomavirus encoded microRNAs. *Plos One, 8*(7), e70202.

Ragland, W. L., Keown, G. H., & Gorham, J. R. (1966). An Epizootic of Equine Sarcoid. *Nature, 210*(5043), 1399-1399.

Ragland, W. L., Keown, G. H., & Spencer, G. R. (1970a). Equine Sarcoid. *Equine Veterinary Journal, 2*(1), 2-11.
Ragland, W. L., McLaughlin, C. A., & Spencer, G. R. (1970b). Attempts to Relate Bovine Papilloma Virus to the Cause of Equine Sarcoid: Horses, Donkeys and Calves Inoculated with Equine Sarcoid Extracts. *Equine Veterinary Journal, 2*(4), 168-172.
Rector, A., & Van Ranst, M. (2013). Animal papillomaviruses. *Virology, 445*(1-2), 213-223.
Reid, S. W., Gettinby, G., Fowler, J. N., & Ikin, P. (1994). Epidemiological observations on sarcoids in a population of donkeys (Equus asinus). *Vet Rec, 134*(9), 207-211.
Reid, S. W., & Mohammed, H. O. (1997). Longitudinal and cross-sectional studies to evaluate the risk of sarcoid associated with castration. *Canadian Journal of Veterinary Research, 61*(2), 89-93.
Roberts, W. D. (1970). Experimental treatment of equine sarcoid. *Vet Med Small Anim Clin, 65*(1), 67-73.
Roden, R. B., Lowy, D. R., & Schiller, J. T. (1997). Papillomavirus is resistant to desiccation. *J Infect Dis, 176*(4), 1076-1079.
Roperto, S., Borzacchiello, G., Esposito, I., Riccardi, M., Urraro, C., Luc‡, R., Corteggio, A., TatË, R., Cermola, M., Paciello, O., & Roperto, F. (2012). Productive Infection of Bovine Papillomavirus Type 2 in the Placenta of Pregnant Cows Affected with Urinary Bladder Tumors. *Plos One, 7*(3), e33569.
Roperto, S., Russo, V., Ozkul, A., Sepici-Dincel, A., Maiolino, P., Borzacchiello, G., Marcus, I., Esposito, I., Riccardi, M. G., & Roperto, F. (2013). Bovine papillomavirus type 2 infects the urinary bladder of water buffalo (Bubalus bubalis) and plays a crucial role in bubaline urothelial carcinogenesis. *J Gen Virol, 94*(Pt 2), 403-408.
Rosenfeld, N., Aharonov, R., Meiri, E., Rosenwald, S., Spector, Y., Zepeniuk, M., Benjamin, H., Shabes, N., Tabak, S., Levy, A., Lebanony, D., Goren, Y., Silberschein, E., Targan, N., Ben-Ari, A., Gilad, S., Sion-Vardy, N., Tobar, A., Feinmesser, M., Kharenko, O., Nativ, O., Nass, D., Perelman, M., Yosepovich, A., Shalmon, B., Polak-Charcon, S., Fridman, E., Avniel, A., Bentwich, I., Bentwich, Z., Cohen, D., Chajut, A., & Barshack, I. (2008). MicroRNAs accurately identify cancer tissue origin. *Nat Biotechnol, 26*(4), 462-469.
Roth, P., Keller, A., Hoheisel, J. D., Codo, P., Bauer, A. S., Backes, C., Leidinger, P., Meese, E., Thiel, E., Korfel, A., & Weller, M. (2015). Differentially regulated miRNAs as prognostic biomarkers in the blood of primary CNS lymphoma patients. *Eur J Cancer, 51*(3), 382-390.
Rothacker, C. C., Boyle, A. G., & Levine, D. G. (2015). Autologous vaccination for the treatment of equine sarcoids: 18 cases (2009-2014). *Can Vet J, 56*(7), 709-714.
Rous, P., & Beard, J. W. (1935). The progression to carcinoma of virus-induced rabbit papillomas. *Journal of Experimental Medicine, 62*, 523-548.
Sanchez-Casanova, R. E., Masri-Daba, M., Alonso-Diaz, M. A., Mendez-Bernal, A., Hernandez-Gil, M., & Fernando-Martinez, J. A. (2014). Prevalence of cutaneous pathological conditions and factors associated with the presence of skin wounds in working equids in tropical regions of Veracruz, Mexico. *Trop Anim Health Prod, 46*(3), 555-561.

Schaffer, P. A., Wobeser, B., Martin, L. E. R., Dennis, M. M., & Duncan, C. G. (2013). Cutaneous neoplastic lesions of equids in the central United States and Canada: 3,351 biopsy specimens from 3,272 equids (2000-2010). *J Am Vet Med Assoc, 242*(1), 99-104.

Schwind, S., Maharry, K., Radmacher, M. D., Mrozek, K., Holland, K. B., Margeson, D., Whitman, S. P., Hickey, C., Becker, H., Metzeler, K. H., Paschka, P., Baldus, C. D., Liu, S., Garzon, R., Powell, B. L., Kolitz, J. E., Carroll, A. J., Caligiuri, M. A., Larson, R. A., Marcucci, G., & Bloomfield, C. D. (2010). Prognostic significance of expression of a single microRNA, miR-181a, in cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol, 28*(36), 5257-5264.

Scott, D., Miller, W., & Griffin, C. (2001). Neoplastic and non-neoplastic tumors. *Muller and Kirk's small animal dermatology, 6*, 1365-1369.

Sharma, G., Dua, P., & Agarwal, S. M. (2014). A Comprehensive Review of Dysregulated miRNAs Involved in Cervical Cancer. *Current Genomics, 15*(4), 310-323.

Shi, M., Du, L., Liu, D., Qian, L., Hu, M., Yu, M., Yang, Z., Zhao, M., Chen, C., & Guo, L. (2012). Glucocorticoid regulation of a novel HPV-E6-p53-miR-145 pathway modulates invasion and therapy resistance of cervical cancer cells. *J Pathol, 228*(2), 148-157.

Shukla, P., Vogl, C., Wallner, B., Rigler, D., Muller, M., & Macho-Maschler, S. (2015). High-throughput mRNA and miRNA profiling of epithelial-mesenchymal transition in MDCK cells. *BMC Genomics, 16*(1), 944.

Silva, M. A., Altamura, G., Corteggio, A., Roperto, F., Bocaneti, F., Velescu, E., Freitas, A. C., Carvalho, C. C. R., Cavalcanti, K. P. S., & Borzacchiello, G. (2013a). Expression of connexin 26 and bovine papillomavirus E5 in cutaneous fibropapillomas of cattle. *Vet J, 195*(3), 337-343.

Silva, M. A., De Albuquerque, B. M., Pontes, N. E., Coutinho, L. C., Leitao, M. C., Reis, M. C., Castro, R. S., & Freitas, A. C. (2013b). Detection and expression of bovine papillomavirus in blood of healthy and papillomatosis-affected cattle. *Genet Mol Res, 12*(3), 3150-3156.

Silvestre, O., Borzacchiello, G., Nava, D., Iovane, G., Russo, V., Vecchio, D., D'Ausilio, F., Gault, E. A., Campo, M. S., & Paciello, O. (2009). Bovine papillomavirus type 1 DNA and E5 oncoprotein expression in water buffalo fibropapillomas. *Vet Pathol, 46*(4), 636-641.

Smith, K. T., Patel, K. R., & Campo, M. S. (1984). Papillomavirus research: a growth area. *Microbiol Sci, 1*(1), 5-8.

Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol, 3*, Article3.

Soon, P., & Kiaris, H. (2013). MicroRNAs in the tumour microenvironment: big role for small players. *Endocrine-related cancer, 20*(5), R257-R267.

Spalholz, B. A., Lambert, P. F., Yee, C. L., & Howley, P. M. (1987). Bovine papillomavirus transcriptional regulation: localization of the E2-responsive elements of the long control region. *J Virol, 61*(7), 2128-2137.

Srivastava, S. K., Bhardwaj, A., Leavesley, S. J., Grizzle, W. E., Singh, S., & Singh, A. P. (2013). MicroRNAs as potential clinical biomarkers: emerging approaches for their detection. *Biotech Histochem, 88*(7), 373-387.

Staiger, E. A., Tseng, C. T., Miller, D., Cassano, J. M., Nasir, L., Garrick, D., Brooks, S. A., & Antczak, D. F. (2016). Host genetic influence on papillomavirus-induced tumors in the horse. *Int J Cancer*.

Stanley, M. (2010). Pathology and epidemiology of HPV infection in females. *Gynecol Oncol, 117*(2 Suppl), S5-10.
Sun, G., Shi, L., Yan, S., Wan, Z., Jiang, N., Fu, L., Li, M., & Guo, J. (2014). MiR-15b targets cyclin D1 to regulate proliferation and apoptosis in glioma cells. *Biomed Res Int, 2014*.
Suzuki, H. I., Yamagata, K., Sugimoto, K., Iwamoto, T., Kato, S., & Miyazono, K. (2009). Modulation of microRNA processing by p53. *Nature, 460*(7254), 529-533.
Szafranski, S. P., Deng, Z.-L., Tomasch, J., Jarek, M., Bhuju, S., Meisinger, C., Kühnisch, J., Sztajer, H., & Wagner-Döbler, I. (2015). Functional biomarkers for chronic periodontitis and insights into the roles of Prevotella nigrescens and Fusobacterium nucleatum; a metatranscriptome analysis. *npj Biofilms and Microbiomes, 1*, 15017.
Tarwid, J. N. F., P.B. and Clark E.G. (1985). Equine sarcoids: A study with emphasis on pathological diagnosis. *The Compendium of Continuing Education, 7*, 293-300.
Taylor, S., & Haldorson, G. (2013). A review of equine sarcoid *Equine Veterinary Education, 25*(4), 210-216.
Teifke, J., Hardt, M., & Weiss, E. (1994). Detection of bovine papillomavirus DNA in formalin-fixed and paraffin-embedded equine sarcoids by polymerse chain reaction and non-radioactive in situ hybridization. *European journal of veterinary pathology: official journal of the European Society of Veterinary Pathology*.
Tong, X., & Howley, P. M. (1997). The bovine papillomavirus E6 oncoprotein interacts with paxillin and disrupts the actin cytoskeleton. *Proceedings of the National Academy of Sciences, 94*(9), 4412-4417.
Torrontegui, B. O., & Reid, S. W. J. (1994). Clinical and pathological epidemiology of the equine sarcoid in a referral population. *Equine Veterinary Education, 6*(2), 85-88.
Trachtenberg, A. J., Robert, J.-H., Abdalla, A. E., Fraser, A., He, S. Y., Lacy, J. N., Rivas-Morello, C., Truong, A., Hardiman, G., & Ohno-Machado, L. (2012). A primer on the current state of microarray technologies. *Next Generation Microarray Bioinformatics: Methods and Protocols*, 3-17.
Tutar, L., Tutar, E., Ozgur, A., & Tutar, Y. (2015). Therapeutic Targeting of microRNAs in Cancer: Future Perspectives. *Drug Dev Res, 76*(7), 382-388.
Valentine, B. A. (2006). Survey of Equine Cutaneous Neoplasia in the Pacific Northwest. *Journal of Veterinary Diagnostic Investigation, 18*(1), 123-126.
Van Doorslaer, K. (2013). Evolution of the papillomaviridae. *Virology, 445*(1-2), 11-20.
Van Doorslaer, K., Tan, Q., Xirasagar, S., Bandaru, S., Gopalan, V., Mohamoud, Y., Huyen, Y., & McBride, A. A. (2013). The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis. *Nucleic Acids Res, 41*(Database issue), D571-578.
van Dyk, E., Oosthuizen, M. C., Bosman, A. M., Nel, P. J., Zimmerman, D., & Venter, E. H. (2009). Detection of bovine papillomavirus DNA in sarcoid-affected and healthy free-roaming zebra (Equus zebra) populations in South Africa. *J Virol Methods, 158*(1-2), 141-151.
van Dyk, E., Bosman, A. M., van Wilpe, E., Williams, J. H., Bengis, R. G., van Heerden, J., & Venter, E. H. (2011). Detection and characterisation of papillomavirus in skin lesions of giraffe and sable antelope in South Africa. *J S Afr Vet Assoc, 82*(2), 80-85.

Venugopal, S. K., Jiang, J., Kim, T. H., Li, Y., Wang, S. S., Torok, N. J., Wu, J., & Zern, M. A. (2010). Liver fibrosis causes downregulation of miRNA-150 and miRNA-194 in hepatic stellate cells, and their overexpression causes decreased stellate cell activation. *Am J Physiol Gastrointest Liver Physiol, 298*(1), G101-106.

Venuti, A., Paolini, F., Nasir, L., Corteggio, A., Roperto, S., Campo, M. S., & Borzacchiello, G. (2011). Papillomavirus E5: the smallest oncoprotein with many functions. *Mol Cancer, 10*, 140.

Vizoso, M., Puig, M., Carmona, F., Maqueda, M., Velásquez, A., Gómez, A., Labernadie, A., Lugo, R., Gabasa, M., Rigat-Brugarolas, L. G., Trepat, X., Ramírez, J., Moran, S., Vidal, E., Reguart, N., Perera, A., Esteller, M., & Alcaraz, J. (2015). Aberrant DNA methylation in non-small cell lung cancer-associated fibroblasts. *Carcinogenesis, 36*(12), 1453-1463.

Vojtechova, Z., Sabol, I., Salakova, M., Smahelova, J., Zavadil, J., Turek, L., Grega, M., Klozar, J., Prochazka, B., & Tachezy, R. (2016). Comparison of the miRNA profiles in HPV-positive and HPV-negative tonsillar tumors and a model system of human keratinocyte clones. *BMC Cancer, 16*, 382.

von Knebel Doeberitz, M. (2002). New markers for cervical dysplasia to visualise the genomic chaos created by aberrant oncogenic papillomavirus infections. *Eur J Cancer, 38*(17), 2229-2242.

Voss, J. L. (1969). Transmission of equine sarcoid. *Am J Vet Res, 30*(2), 183-191.

Wald, A. I., Hoskins, E. E., Wells, S. I., Ferris, R. L., & Khan, S. A. (2011). Alteration of microRNA profiles in squamous cell carcinoma of the head and neck cell lines by human papillomavirus. *Head & neck, 33*(4), 504-512.

Wang, F., Li, B., & Xie, X. (2016). The roles and clinical significance of microRNAs in cervical cancer. *Histol Histopathol, 31*(2), 131-139.

Wang, J., Czech, B., Crunk, A., Wallace, A., Mitreva, M., Hannon, G. J., & Davis, R. E. (2011). Deep small RNA sequencing from the nematode Ascaris reveals conservation, functional diversification, and novel developmental profiles. *Genome Research, 21*(9), 1462-1477.

Wang, M., Wang, W., Zhang, P., Xiao, J., Wang, J., & Huang, C. (2014a). Discrimination of the expression of paralogous microRNA precursors that share the same major mature form. *Plos One, 9*(3), e90591.

Wang, X., Tang, S., Le, S.-Y., Lu, R., Rader, J. S., Meyers, C., & 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., Li, Y., Hafner, M., Banerjee, N. S., Tang, S., Briskin, D., Meyers, C., Chow, L. T., Xie, X., Tuschl, T., & Zheng, Z. M. (2014b). microRNAs are biomarkers of oncogenic human papillomavirus infections. *Proc Natl Acad Sci U S A, 111*(11), 4262-4267.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics, 10*(1), 57-63.

Wang, Z., & Yang, B. (2010). Microarray and Its Variants for miRNA Profiling *MicroRNA Expression Detection Methods* (pp. 67-79): Springer.

Weiss, E. (1974). Tumours of the soft (mesenchymal) tissues. *Bulletin of the World Health Organization, 50*(1-2), 101-110.

Biological agents. Volume 100 B. A review of human carcinogens, 100 C.F.R. (2012).

Wilson, A. D., Armstrong, E. L., Gofton, R. G., Mason, J., De Toit, N., & Day, M. J. (2013). Characterisation of early and late bovine papillomavirus protein expression in equine sarcoids. *Vet Microbiol, 162*(2-4), 369-380.

Winer, J., Jung, C. K., Shackel, I., & Williams, P. M. (1999a). Development and validation of real-time quantitative reverse transcriptase-polymerase chain reaction for monitoring gene expression in cardiac myocytes in vitro. *Anal Biochem, 270*(1), 41-49.

Winer, J., Jung, C. K. S., Shackel, I., & Williams, P. M. (1999b). Development and validation of real-time quantitative reverse transcriptase–polymerase chain reaction for monitoring gene expression in cardiac myocytesin vitro. *Analytical biochemistry, 270*(1), 41-49.

Witten, D., Tibshirani, R., Gu, S. G., Fire, A., & 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.

Wittmann, J., & Jäck, H.-M. (2010). Serum microRNAs as powerful cancer biomarkers. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer, 1806*(2), 200-207.

Wobeser, B. K., Davies, J. L., Hill, J. E., Jackson, M. L., Kidney, B. A., Mayer, M. N., Townsend, H. G., & Allen, A. L. (2010). Epidemiology of equine sarcoids in horses in western Canada. *Can Vet J, 51*(10), 1103-1108.

Wobeser, B. K., Hill, J. E., Jackson, M. L., Kidney, B. A., Mayer, M. N., Townsend, H. G., & Allen, A. L. (2012). Localization of Bovine papillomavirus in equine sarcoids and inflammatory skin conditions of horses using laser microdissection and two forms of DNA amplification. *J Vet Diagn Invest, 24*(1), 32-41.

Wu, J., Bao, J., Kim, M., Yuan, S., Tang, C., Zheng, H., Mastick, G. S., Xu, C., & Yan, W. (2014). Two miRNA clusters, miR-34b/c and miR-449, are essential for normal brain development, motile ciliogenesis, and spermatogenesis. *Proc Natl Acad Sci U S A, 111*(28), E2851-2857.

Xie, W., Li, M., Xu, N., Lv, Q., Huang, N., He, J., & Zhang, Y. (2013). MiR-181a regulates inflammation responses in monocytes and macrophages. *PLoS One, 8*(3), e58639.

Xu, J., Zhang, W., Lv, Q., & Zhu, D. (2015). Overexpression of miR-21 promotes the proliferation and migration of cervical cancer cells via the inhibition of PTEN. *Oncol Rep, 33*(6), 3108-3116.

Yan, Y., Cui, H., Jiang, S., Huang, Y., Huang, X., Wei, S., Xu, W., & Qin, Q. (2011). Identification of a novel marine fish virus, Singapore grouper iridovirus-encoded microRNAs expressed in grouper cells by Solexa sequencing. *Plos One, 6*(4), e19148.

Yang, J., Gao, T., Tang, J., Cai, H., Lin, L., & Fu, S. (2013). Loss of microRNA-132 predicts poor prognosis in patients with primary osteosarcoma. *Mol Cell Biochem, 381*(1-2), 9-15.

Yeung, C. A., Tsang, T., Yau, P., & Kwok, 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.

Yu, J., Feng, J., Zhi, X., Tang, J., Li, Z., Xu, Y., Yang, L., Hu, Z., & Xu, Z. (2015). Let-7b inhibits cell proliferation, migration, and invasion through targeting Cthrc1 in gastric cancer. *Tumour Biol, 36*(5), 3221-3229.

Yu, Y., Zhang, Y., & Zhang, S. (2013). MicroRNA–92 regulates cervical tumorigenesis and its expression is upregulated by human

papillomavirus–16 E6 in cervical cancer cells. *Oncology letters, 6*(2), 468-474.
Yuan, Z., Gallagher, A., Gault, E. A., Campo, M. S., & Nasir, L. (2007a). Bovine papillomavirus infection in equine sarcoids and in bovine bladder cancers. *Vet J, 174*(3), 599-604.
Yuan, Z. Q., Gallagher, A., Gault, E. A., Campo, M. S., & Nasir, L. (2007b). Bovine papillomavirus infection in equine sarcoids and in bovine bladder cancers. *Vet J, 174*(3), 599-604.
Yuan, Z. Q., Philbey, A. W., Gault, E. A., Campo, M. S., & Nasir, L. (2007c). Detection of bovine papillomavirus type 1 genomes and viral gene expression in equine inflammatory skin conditions. *Virus Res, 124*(1-2), 245-249.
Yuan, Z. Q., Gault, E. A., Gobeil, P., Nixon, C., Campo, M. S., & Nasir, L. (2008a). Establishment and characterization of equine fibroblast cell lines transformed in vivo and in vitro by BPV-1: model systems for equine sarcoids. *Virology, 373*(2), 352-361.
Yuan, Z. Q., Nicolson, L., Marchetti, B., Gault, E. A., Campo, M. S., & Nasir, L. (2008b). Transcriptional changes induced by bovine papillomavirus type 1 in equine fibroblasts. *J Virol, 82*(13), 6481-6491.
Yuan, Z. Q., Bennett, L., Campo, M. S., & Nasir, L. (2010a). Bovine papillomavirus type 1 E2 and E7 proteins down-regulate Toll Like Receptor 4 (TLR4) expression in equine fibroblasts. *Virus Res, 149*(1), 124-127.
Yuan, Z. Q., Gobeil, P. A., Campo, M. S., & Nasir, L. (2010b). Equine sarcoid fibroblasts over-express matrix metalloproteinases and are invasive. *Virology, 396*(1), 143-151.
Yuan, Z. Q., Gault, E. A., Campo, M. S., & Nasir, L. (2011a). Upregulation of equine matrix metalloproteinase 1 by bovine papillomavirus type 1 is through the transcription factor activator protein-1. *Journal of general virology, 92*(11), 2608-2619.
Yuan, Z. Q., Gault, E. A., Campo, M. S., & Nasir, L. (2011b). Different contribution of bovine papillomavirus type 1 oncoproteins to the transformation of equine fibroblasts. *J Gen Virol, 92*(Pt 4), 773-783.
Yuan, Z. Q., Gault, E. A., Campo, M. S., & Nasir, L. (2011c). p38 mitogen-activated protein kinase is crucial for bovine papillomavirus type-1 transformation of equine fibroblasts. *J Gen Virol, 92*(Pt 8), 1778-1786.
Zhang, G., Liu, Z., Xu, H., & Yang, Q. (2016). miR-409-3p suppresses breast cancer cell growth and invasion by targeting Akt1. *Biochemical and biophysical research communications, 469*(2), 189-195.
Zhang, H., Hao, Y., Yang, J., Zhou, Y., Li, J., Yin, S., Sun, C., Ma, M., Huang, Y., & Xi, J. J. (2011a). Genome-wide functional screening of miR-23b as a pleiotropic modulator suppressing cancer metastasis. *Nat Commun, 2*, 554.
Zhang, J., Li, S., Yan, Q., Chen, X., Yang, Y., Liu, X., & Wan, X. (2013). Interferon-beta induced microRNA-129-5p down-regulates HPV-18 E6 and E7 viral gene expression by targeting SP1 in cervical cancer cells. *Plos One, 8*(12), e81366.
Zhang, S., Hao, J., Xie, F., Hu, X., Liu, C., Tong, J., Zhou, J., Wu, J., & Shao, C. (2011b). Downregulation of miR-132 by promoter methylation contributes to pancreatic cancer development. *Carcinogenesis, 32*(8), 1183-1189.
Zheng, Z. M., & Wang, X. (2011). Regulation of cellular miRNA expression by human papillomaviruses. *Biochim Biophys Acta, 1809*(11-12), 668-677.

Zhou, M., Wang, Q., Sun, J., Li, X., Xu, L., Yang, H., Shi, H., Ning, S., Chen, L., Li, Y., He, T., & Zheng, Y. (2009). In silico detection and characteristics of novel microRNA genes in the Equus caballus genome using an integrated ab initio and comparative genomic approach. *Genomics, 94*(2), 125-131.

Zhu, S., Pan, W., Song, X., Liu, Y., Shao, X., Tang, Y., Liang, D., He, D., Wang, H., Liu, W., Shi, Y., Harley, J. B., Shen, N., & Qian, Y. (2012). The microRNA miR-23b suppresses IL-17-associated autoimmune inflammation by targeting TAB2, TAB3 and IKK-alpha. *Nat Med, 18*(7), 1077-1086.

Zimmerman, D. (2004). *Summary report and draft action plan for the equine sarcoid problem in cape mountain zebra at Bontebok NP*.In: Internal Report, November 2004, Veterinary Wildlife Unit, South African National Parks Board. Kimberley

Zimmermann, H., Koh, C. H., Degenkolbe, R., O'Connor, M. J., Müller, A., Steger, G., Chen, J. J., Lui, Y., Androphy, E., & Bernard, H. U. (2000). Interaction with CBP/p300 enables the bovine papillomavirus type 1 E6 oncoprotein to downregulate CBP/p300-mediated transactivation by p53. *J Gen Virol, 81*(Pt 11), 2617-2623.